2011

Detection and Genotyping of Cryptosporidium Oocysts in Eastern Pennsylvania Water Supplies

Colin Michael McLeod
Lehigh University

Detection and Genotyping of *Cryptosporidium* Oocysts in Eastern Pennsylvania Water Supplies

by

Colin Michael McLeod

A Thesis
Presented to the Graduate Research Committee
of Lehigh University
in Candidacy for the Degree of
Master of Science
in
Environmental Engineering

in

Department of Civil and Environmental Engineering

Lehigh University

September 2011

Certificate of Approval

This thesis is accepted and approved in partial fulfillment of the requirements for the degree of Master of Science in Environmental Engineering.

_____

Date

_____

Thesis Advisor

_____

Chairperson of Department

Acknowledgements

I would like to acknowledge my advisor Dr. Kristen Jellison for her guidance and assistance throughout the duration of my project. I would also like to acknowledge Dr. Derek Brown and Dr Arup SenGupta for inspiring me intellectually throughout the course of my Masters studies.

Additionally, I would like to thank the Philadelphia Water Department's Office of Watersheds source water program for assisting with sample collection.

Finally, I'd like to thank all of the graduate and undergraduate students in the Jellison Lab, especially Julie Napotnik for her training and Robin Barnes-Pohjonen whose work with FISH was instrumental to the success of my project.

Table of Contents

List of Tables

List of Figures

Abstract

Water samples were obtained from the Monocacy Creek Watershed and the Schuylkill River Watershed in eastern Pennsylvania in order to monitor for multiple species of *Cryptosporidium* oocysts. Filter samples were collected from the intake at the Philadelphia Water Department's Queen Lane Water Treatment Plant (WTP) in the Schuylkill River in Philadelphia, PA and from Monocacy Creek in Bethlehem, PA. Water filtration was followed by immunomagnetic separation to isolate oocysts and then the oocysts were processed by genotyping (i.e., DNA extraction, nested polymerase chain reaction, cloning and sequencing) to determine the individual species of *Cryptosporidium* to assess the potential threat to human health.

Out of 33 filter samples at the Queen Lane WTP over a period of nine months, six (18.2%) were positive for *Cryptosporidium* oocysts. The phylogenetic analysis of oocyst genotypes showed that five different genotypes were found. The detection of human infectious genotypes in the Schuylkill River Watershed confirmed a potential risk to human health associated with using the Schuylkill River as a drinking water source. Out of 14 filter samples at Monocacy Creek over a period of nine months, one (7.1%) was positive for *Cryptosporidium*.

In addition, a method for genotyping oocysts that have already been processed by fluorescent *insitu* hybridization (FISH) was developed. This method was effective in confirming the presence of human infectious genotypes of *Cryptosporidium* at the Queen Lane WTP intake and at two additional sites.

Introduction

What is *Cryptosporidium*?

Cryptosporidium is a protozoan parasite that causes cryptosporidiosis, a significant diarrheal illness that can occur in both healthy and immunocompromised individuals. *Cryptosporidium* oocysts originate from a variety of sources including agricultural runoff (livestock), wild animals, domestic animals and human sewage or wastewater treatment plant (WWTP) effluent. The oocysts shed in feces and become waterborne, which can result in the contamination of food and water.

Cryptosporidium Lifecycle

Cryptosporidium has a complex, monoxenous lifecycle, involving sexual and asexual reproduction which allows the organism to rapidly multiply within a host, even if only a few oocysts are ingested. *Cryptosporidium* form both thin-walled oocysts, which may excyst within the host to start the auto-infectious cycle, and thick-walled oocysts, which are excreted into the environment. The ingestion of oocysts via fecally-contaminated food and water may be followed by a massive shedding of infective oocysts in feces which reintroduces them into the environment. As many as 10^9 oocysts can be excreted daily from a human host, and domestic and wild animals can excrete oocysts in numbers around the same order of magnitude (Blewett, 1989). After excretion, oocysts can be transported into surface and ground water which may include resources for the public water supply. The oocysts can survive in the environment for months depending on environmental conditions (Robertson et al., 1992). Additionally, *Cryptosporidium* can be transmitted from humans to animals and between different animals,

allowing animals to serve as reservoirs for potentially human-infectious genotypes (Chalmers & Giles, 2010).

Why is *Cryptosporidium* an Issue?

In the fecal-orally transmitted parasitic protozoa reference group, *Cryptosporidium* is the most persistent in the environment, the most resistant to chemical disinfection, and the smallest in size (WHO, 2009). Therefore, *Cryptosporidium* oocysts are difficult to remove by filtration and are not completely removed by conventional wastewater and drinking water treatment methods, making wastewater treatment plants a source of oocyst entry into water networks (Pouillot et al., 2004). *Cryptosporidium's* insensitivity to anticoccidial agents and resistance to chemical disinfection further enhance their persistence in drinking water networks (Thompson et al., 2005). This persistence makes finished drinking water one of the main fecal-oral transmission routes for human infection.

Infection of the host is confined to the apical region of the epithelial cells where the oocysts' auto-infectious cycle interferes with fluid and nutrient adsorption. Infection is particularly problematic in immunocompromised individuals where the illness may lead to death (WHO, 2009). In healthy individuals, *Cryptosporidium* infection can be asymptomatic or can cause self-limited diarrhea (Hoxie et al., 1997). The human health risk is compounded because no definitive cure for cryptosporidiosis exists. Current drug treatment is limited to the prescription, anti-diarrheal medicine, nitazoxanide which has not been proven to be effective in immunosuppressed individuals ("CDC - Cryptosporidiosis - Treatment", 2010).

Cryptosporidium Genotypes

The genus *Cryptosporidium* has at least 21 different recognized species. Greater than 60 genotypes of uncertain status have been identified based on oocyst morphology, infection site, preferential host and genetic specificity (Shi et al., 2010). Previously, the species descriptions were based on morphology and host specificity, but molecular markers like the 18S rRNA gene have uncovered greater complexities in the *Cryptosporidium* taxonomy. Previous investigations have studied *Cryptosporidium* in human feces and found that human infection is predominantly due to the bovine and human genotypes of *C. parvum,* but there are a number of additional genotypes that have been related to human infection (Guyot et al., 2001 and Xiao, 2004). *C. meleagridis* , *C. felis*, *C. canis*, *C. suis*, *C. muris*, *C. andersoni*, *C. hominis* monkey genotype, cervine genotype, and the chipmunk genotype I have also been detected in humans (Robinson et al., 2008). Overall, *C. hominis* and *C. parvum* account for the vast majority of human infections (Cacciò, 2005).

Risk to Humans

The human infectious genotypes of *Cryptosporidium* are responsible for 250 to 500 million infections annually in Asia, Africa, and Latin America (Current & Garcia, 1991). *Cryptosporidium* outbreaks are also a concern in developed areas of the world. In the United States, and estimated 748,000 cases of cryptosporidiosis occur each year (Scallan et al., 2011). A massive outbreak in Milwaukee, Wisconsin in 1993 has increased public awareness of cryptosporidiosis. During the outbreak, which was determined to have been caused by *C. hominis* and *C. parvum* (Zhou et al., 2003), 403,000 people had watery diarrhea attributable to *Cryptosporidium* oocysts and at least 54 deaths were the result of *Cryptosporidium* infection (Hoxie et al., 1997). Water-quality measurements at the Milwaukee Water Works plants were

within required limits during the outbreak and there was no evident mechanical breakdown of its flocculators or filters (MacKenzie et al., 1994). The failure to detect a spike in pathogens in the water supply led to many questions regarding the federal turbidity standards that were supposed to ensure protection against parasites like *Cryptosporidium*. Ultimately, federal agencies in the United States agreed upon the Long Term 2 Enhanced Surface Water Treatment Rule (LT2) (EPA 815-R06-005). The LT2 rule requires two years of direct monthly sampling for *Cryptosporidium* which classify water systems into one of four treatment categories. The higher categories require additional water treatment to reduce *Cryptosporidium* levels by 1.0 to 3.0 log, depending on the category and the treatment mechanisms already in place. The LT2 rule has helped increase the awareness of *Cryptosporidium* as a potentially human-infectious parasite and it has led to a greater understanding of the *Cryptosporidium* populations that inhabit drinking water sources.

Research Objective

The objective of this research project was to sample at water sources in eastern Pennsylvania to expand the current inventory of *Cryptosporidium* in the area. Goals of the investigation included determining (i) if *Cryptosporidium* genotypes found at the Queen Lane Water Treatment Plant (WTP) in Philadelphia (monitored and controlled by the Philadelphia Water Department (PWD)) pose a risk to human health and (ii) possible sources of contamination in the Schuylkill River. The Queen Lane WTP is further downstream from agricultural sources than from municipal/industrial point sources and urban runoff, so it was hypothesized that the genotypes of *Cryptosporidium* found at the Queen Lane sampling location should reflect anthropogenic sources. In addition, the frequency of detection at the Queen Lane WTP intake was compared to the frequency of detection at a sampling site in Monocacy Creek in

Bethlehem, which is not directly impacted by municipal point sources, including wastewater treatment plant (WWTP) discharge. The most likely source of *Cryptosporidium* at the Monocacy Creek sampling location is urban runoff, and urban sources were therefore hypothesized to be reflected in the sampling results.

In previous independent studies both a lack of relationship between wet weather events and *Cryptosporidium* detection (Jellison et al., 2009) and a correlation between rainfall and *Cryptosporidium* detection (Curriero et al., 2001) have been found. The same is true with a seasonal correlation. Some studies have found a relationship between seasons and the incidence of *Cryptosporidium* (Montemayor et al., 2005) while others have found no such correlation (Lynch, 2008). Due to these contradictory findings, this study did not consider the relationship of wet weather events to *Cryptosporidium* detection or the seasonal variation in *Cryptosporidium* detection. Therefore, the focus of this project was to determine the *Cryptosporidium* detection frequency and to expand the current catalog of *Cryptosporidium* species and genotypes found in the area without considering turbidity and seasonal variation.

Fluorescent *in situ* Hybridization and Polymerase Chain Reaction

Cryptosporidium is usually dected in surface water using EPA Method 1622/23 (EPA 815-R-05-001). This method uses water filtration and immunomagnetic separation (IMS) to recover and isolate oocysts from surface water sources and then uses an immunofluorescent antibody (IFA) in order to enumerate low levels of oocysts by microscopy. This microscopy-based method does not allow for speciation or genotyping and lacks the ability to determine if an oocyst is viable or nonviable.

These shortcomings in the EPA method have been overcome in research labs by utilizing polymerase chain reaction (PCR) methods for the rapid detection of *Cryptoisporidium* (Johnson et al., 1995) followed by sequencing of the small subunit (SSU) rRNA gene to molecularly characterize the oocysts present in an environmental sample (Xiao et al., 1999). Following IMS, oocyst DNA is extracted, amplified by PCR, sequenced, and genotyped. This molecular characterization detects and differentiates any species or genotype of *Cryptosporidium* in a sample, provides an accurate molecular characterization of the *Cryptosporidium* oocysts and, therefore, differentiates between human-pathogenic *Cryptosporidium* parasites and those that do not infect humans. Unfortunately, this method cannot differentiate viable from nonviable oocysts which is important in determining the risk to human health. Genotyping also lacks the ability to quantify the oocyst concentration in a water source. Others have experimented using quantitative PCR (qPCR) to attempt to quantify the number of oocysts present in a sample (Fontaine & Guillot, 2002; Di Giovanni & LeChevallier, 2005; and Guy et al., 2003), but the sensitivity of nested PCR, which cannot be coupled with qPCR, is more desirable for environmental samples that contain small numbers of oocysts.

Another method uses fluorescent *in situ* hybridization (FISH) to detect *Cryptosporidium* oocysts in environmental samples (Vesey et al., 1998). FISH is desirable because, unlike PCR, it allows for the quantification of oocysts within a sample. FISH is able to detect all species of *Cryptosporidium* and uses a fluorescently labeled oligonucleotide probe (Cry I probe) that allows for differentiation between viable (actively producing rRNA) *C. parvum* and *C. hominis* from all other species. Therefore, FISH provides some molecular specificity, but it is unable to determine the source of the *Cryptosporidium* unless coupled with sequencing.

Incorporation of FISH and PCR into a Single Method

A method that overcomes the shortcomings of FISH and PCR, that can both quantify and genotype *Cryptosporidium* oocysts in environmental samples, is desired. Therefore, in addition to the watershed sampling, this project investigated a combined FISH-PCR protocol. The molecular data that will be obtained by running PCR on samples previously enumerated by FISH will help confirm FISH results that are positive for viable *C. parvum* or *C. hominis* oocysts. It will also be useful to detect potential false negatives that may occur when running FISH.

Oocysts that are processed by FISH should retain their molecular integrity so that they can be processed by PCR if they are rehydrated and removed from the microscope slide (Di Giovanni et al., 2010). Results from other researchers have indicated that single oocysts seeded onto slides have had an approximately 70% to 83% positive detection rate by PCR after removal, depending on the type of slide used (Di Giovanni et al., 2010). This project investigated the optimization of removing low numbers of *Cryptosporidium* oocysts from treated Meriflour slides. The method was tested on slides that were seeded with stock *Cryptosporidium* oocysts, as well as slides that contained environmental samples from multiple sampling locations.

FISH and PCR methods are reliable to a very low detection limit (in fact, PCR requires only a single copy of a target sequence as a template (Li et al., 1990)), but they are dependent upon the efficiency and reliability of the water filtration method used (Method 1622/23) to collect oocysts from source waters. The filtration method only provides a small grab sample that does not reflect changes that occur within the body of water over a period of time. Therefore, a novel sampling method that takes advantage of *in situ* biofilms that has been developed in the Jellison Lab by Elizabeth Wolyniak should be further investigated (Wolyniak,

2010). This biofilm sampling method is noted here because some of the environmental samples that were used to test the FISH-PCR assay were collected by this method.

Materials & Methods

General Experimental Design

Sampling Locations and Sample Collection

Samples were collected between August 2010 and July 2011. Water filter samples were collected from the Queen Lane WTP intake (*n* = 33) which is located on the Schuylkill River just downstream of its confluence with the Wissahickon Creek in Philadelphia, PA. This location lies within the Schuylkill Watershed (Figure 1) but it is heavily impacted by the Wissahickon Watershed (Figure 2). The Wissahickon Creek has an average streamflow of 70 MGD, 18 of which are due to WWTP discharge. Therefore, the average impact of WWTP discharge is almost 26% of the entire base-flow.

Water filter samples were also collected from the Lehigh Watershed within the Lehigh River basin in Monocacy Creek in Bethlehem, PA near its confluence with the Lehigh River (*n* = 14) (Figure 3). Monocacy Creek has an average streamflow of 36 million gallons per day (MGD), 0.51 MGD of which are due to the Bath Boro Sewage Treatment Plant (STP). This is only 1.4% of the total baseline flow.

Environmental samples that were used for the FISH to PCR assay were collected using the *in situ* biofilm holders. They were placed in Sandy Run, which is located within the Wissahickon Watershed, at locations upstream and downstream from the Township of Abington WWTP (Figure 2). Additionally, starting in May 2011, samples were collected upstream from the Bath Boro STP in Monocacy Creek in Bath, PA; in Saucon Creek in Bethlehem, PA; and upstream and downstream from the Bethlehem WWTP which discharges at the confluence of Saucon

Creek and the Lehigh River (Figure 3). Monocacy Creek and Saucon Creek are located within the Lehigh Watershed and are unimpacted by upstream WWTP discharge. The Bethlehem WWTP is also located within the Lehigh Watershed and, similar to the Sandy Run location, allows for a comparison of *Cryptosporidium* concentrations upstream and downstream from a WWTP discharge. Collection at these sites using the *in situ* biofilm method is on-going as part of Robin Barnes-Pohjonen's project.

The Schuylkill, Wissahickon and the Lehigh watersheds are excellent choices for a *Cryptosporidium* study due to their mixtures of urban, suburban, agricultural and rural land use. For a more detailed look at the sampling locations, see Appendix A.

Figure 1. Queen Lane sampling site within the Schuylkill Watershed (courtesy of Philadelphia Water Department).

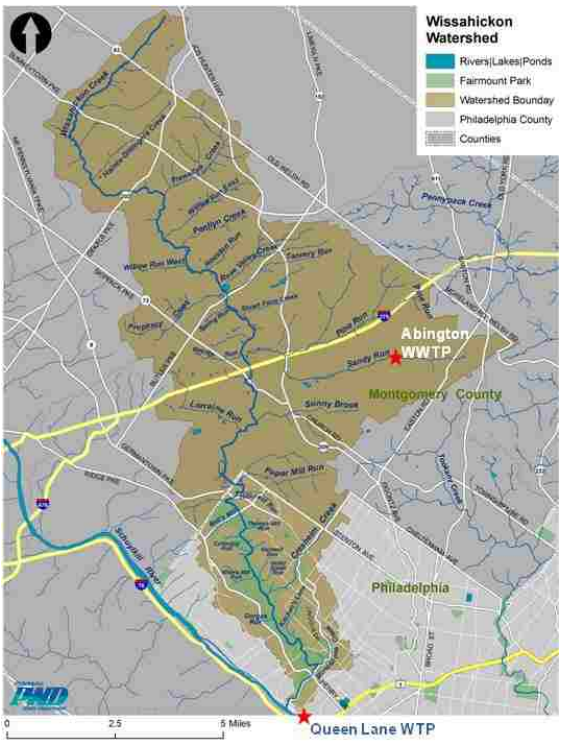

Figure 2. Sampling sites in the Wissahickon Watershed (courtesy of Philadelphia Water Department)

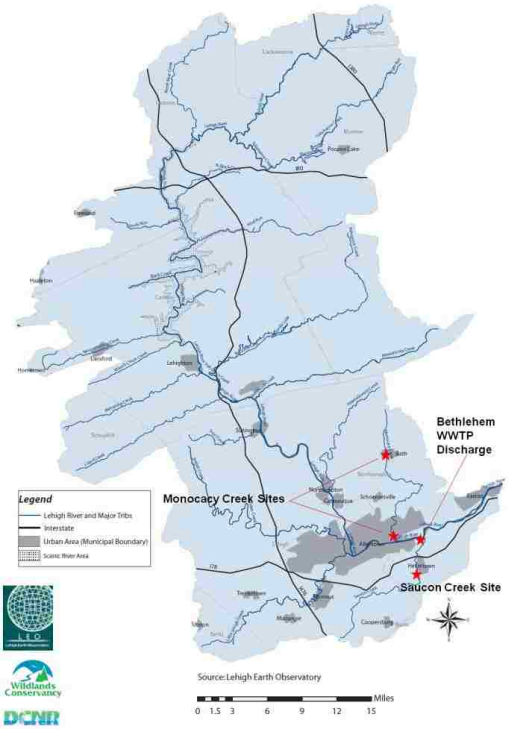

Figure 3. Sampling sites in the Lehigh Watershed (courtesy of Wildlands Conservancy)

Genotyping by PCR

The general experimental design was to process samples obtained by water filtration from Queen Lane and from Monocacy Creek by eluting the filters, followed by IMS, DNA extraction and nested PCR. IMS is a physical separation that takes advantage of *Cryptosporidium*-specific surface antigens that have previously been attached to magnetic beads. Following DNA extraction by a series of density gradients, nested PCR is run using two sets of forward and reverse primers. These primers target a hyper-variable section of the highly conserved small subunit 18S rRNA gene, which is a useful tool for the evolutionary analysis of *Cryptosporidium*. The result is a DNA fragment approximately 434 base pairs in length that can be easily visualized using gel electrophoresis.

In order to determine the speciation of the *Cryptosporidium* in the environmental samples, the secondary PCR products were purified and grown in *E. coli* to select for individual genotypes. Multiple isolates were screened to account for the possibility of the presence of multiple species of *Cryptosporidium* in an individual environmental sample. Finally, sequencing was performed on the individual isolates.

Immunofluorescent Antibody (IFA) and Fluorescent *in situ* Hybridization (FISH)

For samples collected at Sandy Run, Monocacy Creek, Saucon Creek, the Bethlehem WWTP, and some of the samples from Queen Lane, oocysts were identified using a combined IFA and FISH method. These samples were later used to test the FISH to PCR assay. Together, IFA/FISH allows for the detection of oocysts of any species and for the discrimination of viable *C. parvum* and *C. hominis* oocysts from all other species and all non viable oocysts. The IFA method was performed using the Merifluor *Cryptosporidium/Giardia* test kit (Meridian Bioscience, Inc., Cincinnati, OH). The Merifluor kit utilizes a fluorescein isothiocyanite(FITC)-conjugated combination of monoclonal antibodies (mAb) against the cell wall antigens of *Cryptosporidium* and *Giardia*.

FISH employs a fluorescently-labeled oligonucleotide probe , Cry 1 (Sigma-Aldrich Corp., The Woodlands, TX), that targets a specific portion of the 18S rRNA (5'-CGGTTATCCATGTAAGTAAAG-3'). This portion of the rRNA corresponds to the region between nucleotides 138 and 160 of the 18S rRNA gene, a region that is unique to *C. parvum* and *C. hominis*. Because of the short half-life of rRNA, it is only present in high copy numbers in viable organisms (those that continue to make rRNA) (Abelson et al., 1974). Therefore, Cry 1 only stains the viable *C. parvum* and *C. hominis* oocysts.

The mounting media provided in the Merifluor kit contains formalin (formaldehyde). Formalin fixation degrades DNA to the point that the DNA is no longer a suitable template for PCR-based diagnostic tests (Ramos et al., 1999). Therefore, to ensure the success of the FISH to PCR protocol, it was imperative to use a formalin free mounting media (Waterborne Inc., New Orleans, LA).

Experimental Procedures

Water Filtration and Elution

Samples from the Queen Lane WTP intake were filtered with Pall Envirocheck HV Sampling Capsules (Pall Life Sciences, Ann Arbor, MI) according to the manufacturer's recommendations. Sampling was performed in duplicate by PWD personnel twice per month from August 2010 to May 2011 (Table 1). The water filters were shipped overnight to Lehigh University for elution. Monocacy Creek samples were filtered by Lehigh University students once or twice per month from September 2010 to May 2011 (Table 2).

All samples were processed the same day they were collected or received by Lehigh University. The filters were eluted using an elution buffer that consisted of 10 mL of Laureth-12; 10 mL of 1 M Tris pH 7.4; 2 mL of 0.5 M ethylenediamine tetraacetic acid (EDTA) disodiumsalt dihydrate, pH 8.0; and 150 µL of antifoam A, which was brought to a total volume of 1 L using Millipore water (Milli-Q Biocel System; Millipore Corporation, Bedford, MA). The eluted water pellets were resuspended in 5 mL of Millipore water for each 1 mL of pellet.

Water sample volumes varied from 3.8 L to 76.8 L. This variation was due to the propensity of the filter to clog following wet weather events when the turbidity of the stream

was extremely high. The goal was to sample greater than 10 L and this was met for 88% of samples. The pellet size, which is a direct correlation to the amount of suspended solids collected in the filter, varied between less than 1 mL and greater than 3 mL.

Table 1. Summary of sample collection at Queen Lane - PWD sampling site id 5902

Date	Location	Volume Sampled (Liters)	PCR Result Band at ≈434bp	NCBI Blast Result
8/24/2010	Queen Lane	46.7	Negative	
9/21/2010	Queen Lane (A)	Unknown	Negative	
9/21/2010	Queen Lane (B)	Unknown	Inconclusive	
9/28/2010	Queen Lane (A)	65.1	Positive	***C. parvum***
9/28/2010	Queen Lane (B)	71.4	Positive	**Not *Cryptosporidium***
10/12/2010	Queen Lane (A)	42.2	Negative	
10/12/2010	Queen Lane (B)	56.2	Negative	
10/26/2010	Queen Lane (A)	60.9	Negative	
10/26/2010	Queen Lane (B)	69.7	Negative	
11/9/2010	Queen Lane (A)	61.7	Negative	
11/9/2010	Queen Lane (B)	68.5	Negative	
11/23/2010	Queen Lane (A)	68.9	Negative	
11/23/2010	Queen Lane (B)	76.8	Positive	***C. parvum***
12/7/2010	Queen Lane (A)	23.5	Negative	
12/7/2010	Queen Lane (B)	18.9	Negative	
12/21/2010	Queen Lane (A)	24.2	Negative	
12/21/2010	Queen Lane (B)	20.8	Negative	
1/11/2010	Queen Lane (A)	40.9	Positive	***Deer Mouse III genotype***
1/11/2010	Queen Lane (B)	50.3	Positive	***C. suis (Pig I genotype)***
1/25/2011	Queen Lane (A)	54.9	Negative	
1/25/2011	Queen Lane (B)	52.2	Positive	
2/8/2011	Queen Lane (A)	34.1	Negative	
2/8/2011	Queen Lane (B)	26.9	Negative	
2/23/2011	Queen Lane (A)	35.6	Negative	
2/23/2011	Queen Lane (B)	34.8	Positive	***C. andersoni***
3/8/2011	Queen Lane (A)	4.9	Negative	
3/8/2011	Queen Lane (B)	3.8	Negative	
3/22/2011	Queen Lane (A)	25.0	Negative	
3/22/2011	Queen Lane (B)	27.3	Negative	
4/19/2011	Queen Lane (A)	14.8	Negative	
4/19/2011	Queen Lane (B)	12.1	Negative	
4/26/2011	Queen Lane (A)	9.5	Positive	**Not *Cryptosporidium***
4/26/2011	Queen Lane (B)	7.6	Positive	**Not *Cryptosporidium***

Table 2. Sampling at Monocacy Creek. Samples were processed by FISH. The 10/21/2010 sample was lost due to the coverslip falling off the slide and the 3/23/2011 sample was lost between IMS and processing by FISH.

Date	Volume Sampled	Viable Oocysts	Non-Viable Oocysts
9/21/2010	16.55	0	0
10/21/2010	19.11	No data available	
11/8/2010	17.65	0	0
11/23/2010	18.46	0	0
12/8/2010	18.81	0	0
12/22/2010	23.00	0	0
1/11/2011	19.67	0	0
1/25/2011	18.99	0	0
2/9/2011	19.67	0	0
2/24/21011	18.75	0	0
3/9/2011	18.46	0	0
3/23/2011	19.54	No data available	
4/20/2011	21.74	1	0
4/28/2011	18.18	0	0

Biofilm Slides

Biofilms were grown *in situ* for 13 to 19 days on microscope slides placed inside the protective holder which was submerged at the Sandy Run, Monocacy Creek, Saucon Creek, the Bethlehem WWTP, and Queen Lane sampling sites. The slides were scraped clean using a cell scraper and washed with Millipore water. The extract and rinse water was collected and processed by IMS.

IMS

In order to separate the oocysts from the pellet, the 5 mL suspension of Millipore water and water pellet was processed by IMS using the *Cryptosporidium* IMS Beads kit (Virusys Corporation, Taneytown, MD). If the total volume of the pellet exceeded 1 mL, the sample was split into multiple IMS tubes. The manufacturer's recommendations were followed to attach the oocysts to the beads. Oocysts were dissociated from IMS beads using 0.05 M HCl, and the

suspension was neutralized using 0.5 M NaOH. The suspension was centrifuged for 3 minutes at 13,000 rpm, the supernatant was removed and the purified oocysts were resuspended in 50 µL of Millipore water.

Positive and negative IMS controls were processed with each set of samples. Positive IMS controls were made by spiking 4.5 mL of Millipore water with 500 µL of a 10^4 oocysts per mL suspension. Stock oocysts were *C. parvum* Iowa isolates sourced from lab mice (Waterborne Inc., New Orleans, LA) and stored in phosphate buffered saline (PBS) to help maintain a constant pH. Negative IMS controls consisted of 5 mL of Millipore water.

DNA Extraction

Following IMS, the 50 µL IMS product was added to a solution of 450 µL of Tris-EDTA (TE) buffer containing 0.2 g of proteinase K per liter, and 20 µL of 10% sodium dodecyl sulfate (SDS). The oocysts were lysed during an overnight incubation at 45^{o}C.

Positive and negative DNA extraction controls were processed with each set of samples. Positive DNA extraction controls consisted of 50 µL of a 10^4 oocysts per mL stock in 450 µL of TE buffer. Negative DNA extraction controls consisted of 500 µL of TE buffer.

After the incubation, the DNA was extracted twice using phenol: chloroform: isoamyl alcohol (25:24:1 mixture) and once using 99.8% chloroform. The DNA was precipitated with 0.2 M NaCl and one volume of 70% ethanol and another volume of 95% absolute ethanol. After centrifuging twice for 3 minutes at 13000 rpm, discarding the supernatant and allowing the DNA to dry, the DNA was resuspended in 30 µL of TE buffer and allowed to sit for over 6 hours in the refrigerator at 4 ^{o}C.

Nested Polymerase Chain Reaction (PCR)

The hypervariable region of the *Cryptosporidium* 18S rRNA gene was amplified using nested PCR. PCR is very susceptible to contamination with exogenous DNA sequences, and investigators may inadvertently introduce potential target sequences into equipment, solutions and enzymes used in the reaction. Therefore, to ensure sterile conditions, PCR was carried out in a laminar flow hood equipped with UV lights in an area separated from the rest of the lab. In addition, sterile Aerosol Resistant Tips were used, and post amplification processing was completed in a separate room from the PCR area.

The PCR amplification was performed in 50 µL reactions. The primary PCR amplification was performed with 30 µL of DNA template and 20 µL of the PCR cocktail (Appendix B). The secondary PCR amplification was performed using 1 µL of the primary amplification product and 49 µL of the PCR cocktail (Appendix B). The PCR amplifications were completed using *Taq* polymerase and forward and reverse oligonucleotide primers that are complementary to a specific 18S rRNA gene sequence. The forward and reverse primers used in the primary amplification were KLJ1 and KLJ2, respectively (Jellison et al., 2002). The forward and reverse primers used in the secondary amplification were CPB-DIAGF and CPB-DIAGR, respectively (Johnson et al., 1995). The deoxynucleoside triphosphates (Fisher Scientific, Fair Lawn, NJ) were added in equimolar amounts and the cocktail was completed according to the concentrations in Appendix B.

Cycling conditions were previously programmed on the MJ Research PTC-200 Thermocycler (Bio-rad, Hercules, CA). These conditions were determined by previous research (Jellison et al., 2004) and were the same for the primary and the secondary PCRs. Cycling conditions consisted of an initial denaturation (5 minutes at 80 ^{o}C followed by 30 seconds at 98

°C), 25 cycles of amplification (denaturation for 30 seconds at 94 °C, annealing for 30 seconds at 55 °C, and extension for 1 minute at 72 °C), and a final extension (10 minutes at 72 °C).

Positive and negative PCR controls were included with each set of samples. For the primary amplification, positive PCR controls contained 29 µL of sterile Millipore water (Millipore water that has been autoclaved) and 1 µL of genomic *C. parvum* DNA at a DNA concentration equivalent to 10^4 oocysts per µL. This DNA was obtained by running the DNA extraction assay on *C. parvum* Iowa isolate stock oocysts. For the primary amplification, negative PCR controls contained 30 µL of sterile Millipore water. For the secondary amplification, the positive PCR controls contained 1 µL of genomic *C. parvum* DNA at a DNA concentration equivalent to that of 10^4 oocysts per µL. The negative PCR controls contained 1 µL of sterile Millipore water.

After electrophoresis on a 1.4% agarose gel stained with ethidium bromide, secondary PCR products were visualized using UV light. Photographs were obtained after approximately 30 and 60 minutes of electrophoresis and 1 kilobase pair ladders were used as a reference to determine the length of each band (see Appendix C). All primary PCR products were saved and stored at 4 °C in case further processing was required.

Cloning

The secondary PCR products that were positive for *Cryptosporidium* were purified using the QIAquick PCR Purification Kit or the QIAquick Gel Extraction Kit (Qiagen, Inc., Valencia, CA). The purified products were ligated and cloned using the pGEM-T Easy Vector System (Promega Corp, Madison, WI), and used to transform z-competent DH5α *E. coli* cells (Zymo Research, Orange, CA), a strain of *E. coli* that is easily transformed with plasmid DNA. The pGEM-T vector carries the amp^r gene which allows for ampicillin resistance. The *E. coli* were grown on SGAL-

AMP agar so that only the transformed bacteria containing recombinant plasmids (those with ampicillin resistance, the disrupted *lacZ* gene and little to no β-galactosidase activity) could be selected. Twelve clones per PCR product were selected and plated on Luria-Bertani (LB) agar supplemented with 100 µg/mL of ampicillin and single transformed colonies were selected to inoculate LB broth supplemented with 100 µg/mL of ampicillin. After incubating the broth culture overnight, and allowing the bacteria to be grown into a late log phase, the plasmid DNA was isolated from clones by following the manufacturer's recommendations using the QIAPrep Spin Miniprep Kit (Qiagen, Inc., Valencia, CA).

The plasmid DNA was digested with *Not*I to verify the presence of the secondary PCR amplicon insert. The plasmid DNA was also digested with *Nde*I to identify potential heterogeneity among the clones. *Nde*I cuts at the recognition sequence 5'-CAˇTATG-3' 5'-GTATˇAC-3'. This particular restriction recognition site is found in the 18S rRNA gene in *C. hominis, C. parvum, C. suis, C. meleagridis, C. wrairi,* and the deer III mouse genotypes. Restriction enzymes were obtained from Fisher BioReagents (Fair Lawn, NJ). Restriction digestion was carried out in a 20 µL volume containing 4 µL of plasmid DNA, 10 U (1 µL) of *Not*I, 10 U (1 µL) of *Nde*I, 11.8 µL of pure H_2O, 2 µL of Buffer D (1.5 M NaCl, 60 mM Tris-HCl, 60 mM $MgCl_2$, 10 mM dithiothreitol), and 0.1 mg of acetylated bovine serum albumin (BSA) per mL (0.2 µL). This solution was incubated at 37^{o}C for one hour. Digestion products were visualized after 60 minutes of electrophoresis on a 1.4% agarose gel stained with ethidium bromide.

The ligase product was saved and stored in the freezer at -20 ^{o}C in case further processing was necessary.

Sequencing

The densities of the cloned DNA were determined using a GeneQuant spectrophotometer (GE Healthcare Bio-Sciences Corp., Piscataway, NJ). Clones that were representative of *Cryptosporidium* spp. were shipped overnight to the University of Pennsylvania School of Medicine DNA Sequencing Facility (Philadelphia, PA). Sequencing was performed using Applied Biosystems 3730XL (Life Technologies Corporation, Carlsbad, CA) with BigDye Taq FS Terminator V 3.1. For each clone, T7 and SP6 primers were added at the sequencing facility.

If there were multiple *Nde*I digestion patterns, the clones of each digestion pattern were sent and sequenced in triplicate if possible. Both strands of each clone were sequenced for confirmation.

IFA/FISH

Following IMS, the samples collected at Sandy Run, Monocacy Creek, Saucon Creek, the Bethlehem WWTP, and some of the samples from Queen Lane were processed by IFA/FISH. The samples were washed and the oocysts were resuspended in 100 µL PBS. The cell walls were permeabilized using 100 µL of acetone and allowing 15 minutes for incubation. After centrifuging and resuspending the permeablized oocysts in 100 µL PBS, 5 µL of the 1 mMol Cry 1 oligonucleotide probe (5'-CGGTTATCCATGTAAGTAAAG-3') was added, and the sample was allowed to incubate for 1 hour at 48 ^{o}C in a dry bath. The probe was originally synthesized by the DNA Analysis Facility at Johns Hopkins University, Baltimore, MD, in 1.0 µM scale, purified by HPLC, and labeled with a single molecule of fluorochrome, hexachlorinated 6-carboxyfluorescein (HEX). After centrifuging and resuspending in 20 µL of pure water, oocysts were stained on

slides with the Merifluor *Cryptosporidium/Giardia* test kit (Meridian Bioscience Inc., Cincinnati, OH) according to the manufacturer's specifications until April 2011, when a new protocol for staining oocysts in suspension was followed (see Appendix D).

Beginning in March 2011, when the FISH to PCR protocol was proven reliable, only formalin free mounting media was used to mount the coverslip to the Merifluor-stained slides.

Combined FISH and PCR Analysis

Following the FISH protocol, slides were wiped with a Kimwipe saturated with 6-10% bleach followed by a Kimwipe saturated with isopropanol. After cleansing the outside of the slides, contact with the face of the slide was minimized. The clear nail polish that was applied to the coverslip in order to view the slides at 100x microscopy was dissolved using acetone-free nail polish remover. A sterile scalpel was used to slowly and carefully remove the cover slip. Excess mounting media outside of the slide well was removed using a Kimwipe without cross contaminating between wells and without touching the surface of the well.

A volume of 15 µL of Millipore water was added to the slide well, distributed around the surface of the well by gently tilting and rotating the slide, aspirated and transferred into a 1.5 mL microcentrifuge tube. Another volume of 15 µL of Millipore water was added to the center of the slide well, and the surface of the well was thoroughly scraped with a cell scraper. The slide was rotated 90 degrees and scraped again. The Millipore water was then aspirated from the slide and transferred to the microcentrifuge tube. This scraping step was repeated, and finally one more 15 µL wash was added to the well and aspirated. The resulting 60 µL sample was centrifuged briefly before proceeding to DNA extraction.

Slides were kept and stored in the dark at 4 ^{o}C until they were microscopically examined to verify oocyst removal and to ensure uniform scraping.

Analytical Methods

Alignment

For each positive PCR sample, multiple clones were sequenced. If multiple digestion patterns existed after digestion with *Nde*I (see Appendix E) at least one clone for each digestion pattern was sequenced, but the goal was to sequence at least three clones from each positive sample and each different digestion pattern. Due to low concentrations of DNA after cloning, only two clones were sequenced for the 1/11/11 QL filter A sample and the 2/23/11 QL sample. Only one clone was sequenced for the 4/26/11 QL sample and it turned out to be an endosymbiotic diatom, not *Cryptosporidium*. Cloning of the positive FISH to PCR samples revealed the expected digestion patterns and at least three clones were sequenced for each positive FISH to PCR sample.

DNA sequences were compared to the GenBank sequence database using the basic local alignment search tool (BLAST) algorithm. The GenBank sequence database is an annotated collection of publicly available nucleotide sequences that allows for the characterization of specific sequences based on previous submissions to GenBank. GenBank includes all sequence data that is submitted to the database, therefore there are many sequencing results for the same loci. It also provides a way of determining genetic variations between species which can lead to the identification of single nucleotide polymorphisms (SNPs) (Mizrachi, 2002).

All of the DNA sequences were analyzed using BioEdit (version 7.0.5.3) (Hall, 1999), followed by manual alignment editing and submission to the MEGA5 tree-building program (Tamura et al., 2011). When multiple clones from a single sample were sequenced with less than 1% difference, the consensus sequence was derived using the built-in BioEdit function. These consensus sequences were compared to the GenBank sequence database in order to determine the correct species, and the consensus sequences were used in the phylogenetic analysis. This approach acts to limit the errors involved in the phylogenetic analysis. Previous studies aligned *Cryptosporidium* sequences according to secondary sequence structure (Lynch, 2008 and Ziemann, 2006), but with the rapid increase in genetic data, the clustal alignment is preferred to the more time intensive secondary structure alignment.

Phylogenetic Analysis

The phylogenetic analysis used the aligned consensus sequences of the hypervariable region of the *Cryptosporidium* 18S rRNA gene and assumed the sequencing readout is correct and that the sequences were homologous. A neighbor-joining phylogenetic tree was created using MEGA5 (Tamura et al., 2011). The neighbor-joining phylogenetic tree is distance-based, meaning it computes the pairwise distances between sequences and uses the amount of dissimilarity between two aligned sequences to derive trees. These distances were computed using the Kimura 2-parameter method, a simple mathematical method that allows for the estimation of evolutionary distances in terms of the number of nucleotide substitutions (Kimura, 1980).

All positions containing alignment gaps were eliminated only in pairwise sequence comparisons. It was important to include the gaps in our phylogenetic analysis since the

phylogenetic signals contained within the insert/deletion regions are of utmost importance to the determination of the evolutionary relationship between genotypes.

In order to evaluate the evolutionary basis of the *Cryptosporidium* species, it is necessary to use a root organism that is not too far or too close to the ingroup. For this analysis, *Eimeria tenella* (accession number AF026388), another apicomplexan parasite was used to root the tree (Figure 4).

Bootstrap Test and Statistical Support

Bootstrapping refers to a statistical analysis used to determine the accuracy of the phylogenetic tree. A parametric bootstrapping maximum likelihood method with 1000 replicates was used (Felsenstein, 1985). This method generates 1000 new data sets from the original consensus alignments and computes the proportion of times a particular branch appeared in the replicate trees derived from those multiple sets. Bootstrap values greater than 70% correspond to a probability of greater than 95% that the phylogeny has been found (Hillis & Bull, 1993). Values greater than 50% are shown on the tree in Figure 4.

Results, Data Analysis and Discussion

Cryptosporidium Detection in Water Samples

In total, 33 water filter samples were collected from the Queen Lane WTP and 14 water filter samples were collected from Monocacy Creek. These samples were examined for the presence of *Cryptosporidium* oocysts. Although nine samples were successfully cloned, after sequencing only six (18.2% of the total Queen Lane water filter samples) were confirmed to be *Cryptosporidium*. The other three samples appeared to have secondary PCR products that were around 434 basepairs long, but sequencing revealed that the segments were slightly larger. These misreads turned out to be commonly found diatoms (cyclotella and discostella). One sample from the Monocacy Creek (7.1%) water filters tested positive for *Cryptosporidium* using FISH. The difference in detection between Queen Lane and Monocacy Creek was expected as the Schuylkill River and its tributaries (upstream of the Queen Lane intake) are heavily impacted by wastewater discharge whereas Monocacy Creek is virtually unimpacted by potential point sources.

Of the six *Cryptosporidium* sequences detected from the Queen Lane WTP intake, four (66.7%) were identified as potentially human-infectious genotypes. Two were *C. parvum* which indicates a potential public health risk. One was *C. andersoni* and one was *C.* suis, which are potentially infectious because they have been associated with human disease, but they do not pose the same risk as *C. parvum.* The *C. parvum* genotype was found in the samples collected on 9/28/10 and 11/23/10, the *C. andersoni* genotype was found on 2/23/11, and the *C. suis* genotype was found on 1/11/11.

The remaining two genotypes were not identified as being associated with human infection and therefore do not pose a risk to human health. The sequence detected on 1/25/11 was determined to be the goose I genotype of *Cryptosporidium*. On 1/11/11, multiple *Cryptosporidium* species/genotypes were detected. In addition to the *C. suis* genotype mentioned in the previous paragraph, a *Cryptosporidium* sequence most closely related to the deer mouse III genotype was detected. The DNA sequences of the genotypes found in the positive samples can be found in Appendix F.

The sample from 1/11/11 exhibits the possibility of detecting multiple genotypes within a single water filter sample. This finding supports taking the additional step of screening a dozen clones (rather than direct sequencing of the PCR product) in order to determine if one or more genotypes are present in a sample. Even after cloning, there may be under-representations of the diversity within a specific water source. It has been reported that when dealing with high ratios of template mixtures (> 1:10), there is likely a reduction or loss of detection of the less abundant species by RFLP analysis, most likely due to heteroduplex formation in the later cycles of the PCR (Ruecker et al., 2011).

The diversity of the *Cryptosporidium* population based on a single year of sampling shows that multiple sources of *Cryptosporidium* can impact a single sampling location. In the case of the Schuylkill River, suspected sources include wastewater treatment plant discharges, combined sewer overflows, storm water runoff from urban sources and storm water runoff from agricultural sources. Additionally, the genotypes detected at the Queen Lane WTP indicate human, wildlife and agricultural sources of *Cryptosporidium*. For instance, *C. parvum,* which was found in 75% of the Queen Lane samples (when the FISH to PCR results are included), has been

traced to human, wildlife (deer), and agricultural (cattle, sheep, goats, and horses) sources (Xiao et al., 2004).

It is important to note that the type of animal the genotype is named after many not necessarily be the source of the *Cryptosporidium* oocyst. Previous studies have shown that animal genotypes can be found in the fecal matter of animals other than their primary hosts (Lynch, 2008 and Jellison et al., 2009). Additionally, cross transmission of oocysts between multiple hosts in a watershed is very likely given the proximity of deer and geese to areas inhabited by humans, domesticated animals, and other wildlife (Jellison et al., 2009). Therefore, source tracking is inconclusive if it is based solely on the genotype of the detected *Cryptosporidium* oocyst.

Phylogenetic Analysis

Neighbor-joining trees were created to determine the phylogenetic relationship between the oocysts obtained from the water samples. Several distinct taxa of *Cryptosporidium* spp. were found in the water samples (Figure 4). The genotypes obtained from environmental samples are labeled to corresponding to the sampling date and location as given in Table 1 and Table 3. The method of collection (filter or biofilm) is also given on the tree, and if there were multiple digestion patterns during cloning, the specific clone is indicated on the tree. If all of the clones produced the same sequencing result, there is no clone number with the label. GenBank sequences are used in the tree to provide a representative collection of *Cryptosporidium* spp. The GenBank entries are labeled with the species name and the accession number.

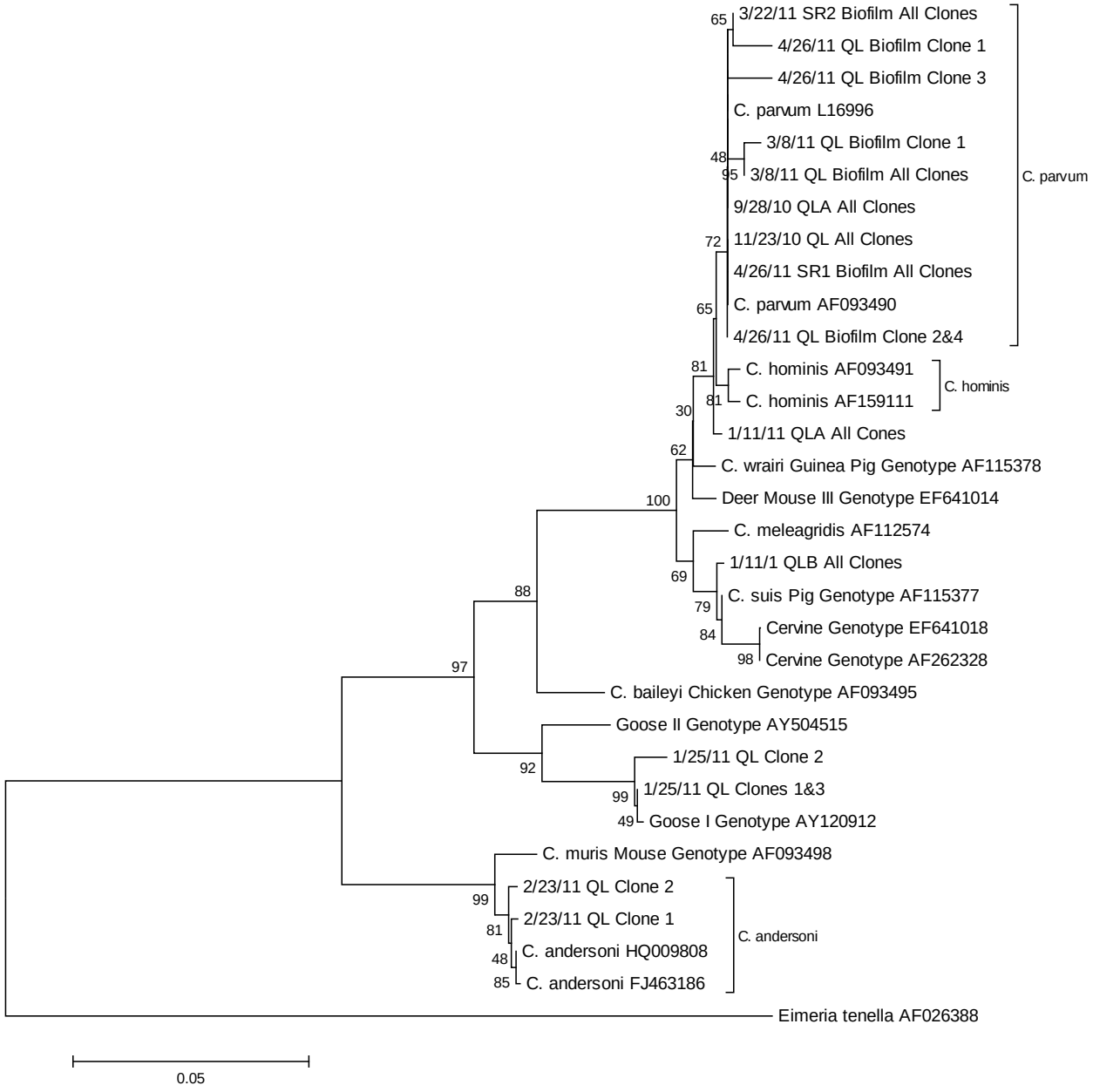

Figure 4. Neighbor joining phylogenetic tree of the samples positive for *Cryptosporidium* that were collected between September 2010 and May 2011.

The phylogenetic tree supports the speciation determined by the Blast search except in the case of the 1/11/11 Queen Lane Filter A sample. The Blast search suggested that this sample was most closely related to the deer mouse III genotype. This could still be the case, but

the phylogenetic tree suggests that it could be more closely related to *C. parvum* or *C. hominis*, or it may even be related to *C. wrairi* or *C. meleagridis.*

The samples found at the Queen Lane intake are clearly separated into three distinct clades. The first and largest contains *C. parvum, C. hominis, C. wrairi, C. meleagridis,* the deer mouse III genotype, *C. suis* and the cervine genotype. The second contains the goose I and goose II genotypes. The third contains *C. andersoni* and *C. muris.* No samples fall into the clade that contains *C.baileyi* alone. The phylogenetic tree shows that every *Cryptosporidium* spp. that was detected using the FISH to PCR method fell into the large clade that represented *C. parvum.*

FISH to PCR Assay Sensitivity Testing

Previous projects have confirmed the sensitivity of the IMS, DNA extraction, and PCR assay as a method for detecting single oocysts from source waters (Jellison et al., 2002 and Lynch, 2008). Therefore, although the sensitivity of the PCR method for molecular characterization of a single oocyst was re-validated for this project, the results will not be presented here.

The sensitivity of the FISH to PCR assay was tested by seeding Merifluor slides with a known quantity of stock oocysts. First, to ensure that the process of removing the oocysts from the Merifluor slides was effective, Merifluor slides were seeded in triplicate with 10^4 stock *C. parvum* oocysts (Waterborne, New Orleans, LA). Formalin-free mounting media was used to fix the coverslip, and the oocysts were processed by IFA/FISH. After counting the oocysts by epifluorescence microscopy, oocysts were removed from the slides. The suspension that was aspirated from the slides was further processed by genomic DNA extraction and nested PCR.

Electrophoresis was run on the secondary PCR product on a 1.4% agarose gel, and the results can be seen in Figure 5.

Next, the detection limit of the FISH to PCR assay was tested using dilutions of stock oocysts. Dilutions were made of 10^4, 10^3, 100, 10 and 1 oocyst(s). The dilutions were seeded onto Merifluor slides, and the oocysts were counted, removed and run through nested PCR. The results show that it is possible to detect a single oocyst (Figure 6). Subsequent runs of the entire assay confirmed the detection of a single oocyst (Figures 7 and 8), but also showed some inconsistency detecting lower oocyst concentrations. In Figure 7 there was no detection of the 10 oocyst dilution, but there was detection of the 1 oocyst dilution. In Figure 8 there was no detection of the 1 oocyst dilution. Therefore, based on the sensitivity testing, it can be suggested that the detection limit lies somewhere below 100 oocysts and further testing should be done with 1, 5, 10, 15, 20, etc. oocysts to determine an accurate detection limit.

An interesting finding is that on all three occasions when 1 oocyst was detected by PCR, no oocysts were detected when counting with FISH. The detection by PCR, but not by FISH, could be the result of a failure of the IFA/FISH process to detect a single oocyst, an inaccurate dilution of a single oocyst, or operator subjectivity when counting the oocysts on the Merifluor slides. This result could also be due to the PCR method having a higher sensitivity than the IFA/FISH method.

The standard error and operator subjectivity associated with the FISH protocol, combined with the standard error associated with stock dilutions suggests that further sensitivity testing should be performed. This sensitivity testing should utilize flow cytometry to seed the Merifluor slides with known oocyst quantities in order to determine a more accurate detection limit associated with the removal of oocysts from the Merifluor slides.

Figure 5. The FISH to PCR assay was first tested by seeding FISH slides with 10^4 oocysts. The secondary PCR products are shown after electrophoresis on a 1.4% agarose gel stained with ethidium bromide after 60 minutes. From left to right, the lanes are as follows: (1) 1kbp ladder, (2-4) samples spiked with 10^4, (5) DNA extraction positive control, (6) DNA extraction negative control, (7) primary PCR positive control, (8) primary PCR negative control, (9) secondary PCR positive control and (10) an additional 1kbp ladder (secondary PCR negative control, which came out as expected, is not shown).

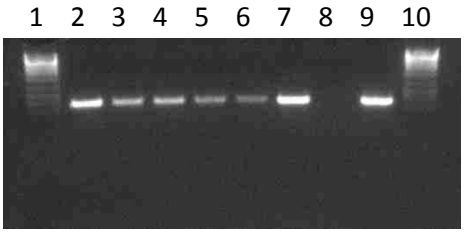

Figure 6. Test of the FISH to PCR assay detection limit. Secondary PCR products are shown after electrophoresis on a 1.4% agarose gel stained with ethidium bromide after 30 minutes. From left to right, the lanes are as follows: (1) 1kbp ladder, (2-6) FISH slides spiked with 10^4, 10^3, 100, 10, and 1 oocyst(s) respectively, (7) DNA extraction positive control, (8) DNA extraction negative control, (9) primary PCR positive control, and (10) an additional 1kbp ladder (primary PCR negative control, secondary PCR positive control and secondary PCR negative control, all of which came out as expected, are not shown).

Figure 7. An additional test of the FISH to PCR assay detection limit. From left to right, the lanes are as follows: (1) 1kbp ladder, (2-6) FISH slides seeded with 1, 10, 100, 10^3, and 10^4 oocyst(s), respectively, (7) DNA extraction positive control, (8) DNA extraction negative control, (9) primary PCR positive control, and (10)an additional 1kbp ladder (primary PCR negative control, secondary PCR positive control and secondary PCR negative control, all of which came out as

expected, are not shown). Notice the detection of one oocyst, as indicated by the faint band in lane 2, but no detection of 10 oocysts in lane 3.

Figure 8. Two additional tests of the FISH to PCR assay detection limit. Gel imaged at 45 minutes. On row 1, from left to right, the lanes are as follows: (1) 1kbp ladder, (2-6) FISH slides seeded with 10^4, 10^3, 100, 10, and 1 oocyst(s), respectively, (7-9) FISH slides seeded with 10^4, 10^3, and 100 oocysts, respectively, and (10) an additional 1kbp ladder. On row 2, from left to right, the lanes are as follows: (10) 1kbp ladder, (12-13) FISH slides seeded 10 and 1 oocyst(s), respectively, (14) DNA extraction positive control, (15) DNA extraction negative control, (16) primary PCR positive control, (17) primary PCR negative control, (18) secondary PCR positive control, (19) secondary PCR negative control, and (20) an additional 1kbp ladder. Notice detection down to 1 oocyst for the first set of dilutions and down to 10 oocysts for the second set. The PCR positive controls did not work due to the use of a DNA stock that did not have a high enough concentration of DNA.

FISH to PCR Assay Results

Following the validation of the FISH to PCR assay, the method was applied to archived PWD samples that were previously processed by IMS and IFA/FISH. The PCR assay was performed on Merifluor slides dating back to 3/8/11, when the formalin free mounting media was first used. In total, the PCR assay was performed on 30 archived samples. These included samples collected by water filtration, samples collected using the *in situ* biofilm method, and samples that were collected by scraping the natural biofilm from the surface of rocks (Table 3). All of the samples were processed by IMS and then counted using IFA/FISH. The samples were stored at 4 ^{o}C until the slide removal was performed.

Table 3. Results from the FISH to PCR assay on archived samples.

Sampling Date	Location	Method	FISH Counts (Viable)	Date of Slide Removal	Number of Days in Storage	PCR
3/8/2011	Queen Lane	Biofilm	8 (4)	7/5/2011	119	Positive
3/8/2011	Sandy Run Up	Biofilm	0 (0)	7/5/2011	119	Negative
3/8/2011	Sandy Run Down	Biofilm	1 (0)	7/5/2011	119	Negative
3/9/2011	Monocacy Creek	Filter	0 (0)	7/5/2011	118	Negative
3/22/2011	Queen Lane	Biofilm	5 (2)	6/9/2011	79	Negative
3/22/2011	Sandy Run Up	Biofilm	1 (0)	6/9/2011	79	Negative
3/22/2011	Sandy Run Down	Biofilm	1 (0)	6/9/2011	79	Positive
3/23/2011	Monocacy Creek	Biofilm	1 (0)	7/5/2011	104	Negative
4/12/2011	Queen Lane	Biofilm	3 (1)	6/9/2011	58	Negative
4/12/2011	Sandy Run Up	Biofilm	0 (0)	6/9/2011	58	Negative
4/12/2011	Sandy Run Down	Biofilm	1 (1)	6/9/2011	58	Negative
4/13/2011	Monocacy Creek	Biofilm	1 (1)	7/5/2011	83	Inconclusive
4/13/2011	Monocacy Creek	Rock Scraping	1 (1)	7/5/2011	83	Negative
4/20/2011	Monocacy Creek	Filter	1 (1)	7/5/2011	76	Negative
4/28/2011	Monocacy Creek	Filter	0 (0)	7/5/2011	68	Negative
4/26/2011	Queen Lane	Biofilm	2 (1)	6/3/2011	38	Positive
4/26/2011	Sandy Run Up	Biofilm	0 (0)	6/3/2011	38	Positive
4/26/2011	Sandy Run Down	Biofilm	0 (0)	6/3/2011	38	Negative
5/10/2011	Queen Lane	Biofilm	0 (0)	6/3/2011	24	Negative
5/10/2011	Sandy Run Up	Biofilm	0 (0)	6/3/2011	24	Negative
5/10/2011	Sandy Run Down	Biofilm	0 (0)	6/3/2011	24	Negative
6/20/2011	Monocacy Creek Up	Biofilm	0 (0)	7/5/2011	15	Negative
6/20/2011	Saucon Creek	Biofilm	2 (0)	7/5/2011	15	Negative
6/22/2011	Bethlehem WWTP Up	Biofilm	0 (0)	7/5/2011	13	Negative
6/22/2011	Bethlehem WWTP Down	Biofilm	1 (0)	7/5/2011	13	Negative
6/29/2011	Monocacy Creek Up	Biofilm	0 (0)	7/12/2011	13	Negative
6/29/2011	Saucon Creek	Biofilm	1 (1)	7/12/2011	13	Negative
7/6/2011	Bethlehem WWTP Up	Biofilm	2 (0)	7/12/2011	6	Negative
7/13/2011	Monocacy Creek Up	Biofilm	0 (0)	7/19/2011	6	Negative
7/13/2011	Saucon Creek	Biofilm	2 (1)	7/19/2011	6	Negative

The results from the removal, combined with the FISH counts, as determined by graduate student Robin Barnes-Pohjonen, can be seen in Table 3. The FISH counts are given as total oocysts and viable oocysts. Viable oocysts are those that have stained with the Cry 1 probe. Of the samples processed with the FISH to PCR assay (*n* = 29), 13 (45%) were positive by FISH but negative by PCR (the inconclusive result from 4/13/11 is not included). One (3%) of the samples processed with the FISH to PCR assay was positive by PCR but negative by FISH. It is possible that this was the result of contamination, and therefore it was removed from subsequent analysis. Twelve (41%) were negative by both FISH and PCR, and three (10%) samples were positive by both FISH and PCR. After eliminating the potentially contaminated sample (5/1/11 Sandy Run Up) from the analysis, the results for FISH and PCR were in agreement for 15 out of 28 samples (54%).

Of the 13 samples that were positive by FISH but negative by PCR, one or two oocysts were counted by FISH 85% of the time. The Shapiro-Wilk test of normality confirms that the values of FISH counts versus positive/negative PCR results after removal from the slides are normally distributed (Appendix G) (Shapiro & Wilk, 1965). After calculating the variances of the FISH counts of the positive samples after removal (FISH+/PCR+, *n* = 3) (the sample from 4/26/11 was removed from the analysis due to potential contamination) and those that were negative after removal (FISH+/PCR-, *n* = 13) (the inconclusive result from 4/13/11 is removed from the analysis), the variance ratio test was performed and compared to the calculated *F*-distribution. The variance ratio between the numbers of oocysts that led to FISH+/PCR+ and those that led to FISH+/PCR- was greater than the *F*-distribution, meaning the two populations cannot be compared using the student *t* test (Appendix G).

The hypothesis that the median number of oocysts detected by FISH in the FISH+/PCR- samples was less than the median number of oocysts detected by FISH in the FISH+/PCR+ was tested using the Mann-Whitney test (see Appendix G) (Mann & Whitney, 1947). The test showed no statistical difference between the number of oocysts detected by FISH, in samples that were positive for PCR versus those that were negative by PCR (p = 0.44). This lack of relationship may be attributable to the small number of FISH+/PCR+ samples in the analysis.

The mean number of oocysts on the Merifluor slides that were positive by FISH and negative by PCR after removal was 1.69, and the mean number of oocysts on the Merifluor slides that were positive by FISH and positive by PCR after removal was 3.67 (Table 4). As discussed earlier, the difference in these two sample populations is not statistically significant. Therefore, the detection limit needs to be further supported by sensitivity testing using flow cytometry to seed Merifluor slides and by seeding samples that contain an environmental matrix.

Table 4. Statistical data associated with the samples that had positive FISH counts and were processed by the FISH to PCR assay. *n* = number of samples, df = degrees of freedom, Mean = average number of oocysts per slide.

	Negative by PCR	Positive by PCR
n	13	3
df	12	2
Mean	1.69	3.67
Variance	1.40	14.33
Std dev	1.18	3.79

There is a possibility that the samples that were positive by FISH but negative by PCR were stored for too long to get accurate results. DNA can degrade over time unless frozen, and the Merifluor slides were only stored at 4 ^{o}C. In order to determine if the storage time had an

impact on the detection by PCR, the positive PCR samples were compared to the negative PCR samples on the basis of the number of days between sample collection and the removal of the oocysts from the Merifluor slides. Both sets of data were normally distributed and the variances were similar (see Table 5). The student's *t* test was used to test the significance (Appendix G). The test statistic (t* = 0.862) is less than the critical *t* value at 95% confidence ($t^{c}_{28, 0.05}$ = 1.7011). Therefore the null hypothesis, that there is no difference between the populations, stands and it can be inferred that there is no difference between the storage times of FISH slides which tested positive versus negative by PCR.

Table 5. The statistical data associated with the samples that had positive FISH counts and were processed by the FISH to PCR assay based on the number of days between sampling and removal of the oocysts from the slides. *n* = number of samples, df = degrees of freedom, Mean = average number of days between sampling and removal.

	Negative by PCR	Positive by PCR
n	26	3
df	25	2
Mean	50.46	78.67
Variance	1520.58	1640.33
Std dev	38.99	40.50

In the case of samples that were positive by FISH but negative by PCR and the case of the sample that was negative by FISH but positive by PCR, it is possible that the inconsistencies were the result of operator subjectivity with respect to the FISH counting. Finally, although not seen in this study, the increased sensitivity of PCR compared to FISH could also lead to inconsistencies in the results from the FISH to PCR data (Morgan & Thompson, 1998).

Sequencing of Environmental Samples from Queen Lane and Sandy Run

Sequencing data was obtained from the samples that were positive by PCR after the slide removal assay (3/8/11 Queen Lane, 3/22/11 Sandy Run Downstream, 4/26/11 Queen Lane, 4/26/11 Sandy Run Downstream). These samples were ligated, cloned, and sent to UPenn for sequencing. The results are shown in Table 6.

Table 6. Sequencing results from removal of oocysts from Merifluor slides following FISH. 4/26/11 SR Upstream was potentially contaminated.

Sampling Date	Location	FISH Counts (Total)	FISH Counts (Viable)	Sequencing Result
3/8/2011	Queen Lane	8	4	*C. parvum*
3/22/2011	Sandy Run Downstream	1	0	*C. parvum*
4/26/2011	Sandy Run Upstream	0	0	*C. parvum*
4/26/2011	Queen Lane	2	1	*C. parvum*

These data help support the water filtration sequencing data obtained from Queen Lane. Both of the Queen Lane biofilm samples that were processed by the FISH to PCR assay were determined to be *C. parvum*. Accounting for the data from sequencing the oocysts on the biofilm slides, brings the percentage of human infectious genotypes of *Cryptosporidium* found at the Queen Lane WTP between September 2010 and May 2011 to 75% of positive samples (67% of which were *C. parvum*). This information further supports the finding that there is a potential risk to human health at the Queen Lane WTP intake.

The result from the 3/22/11 Sandy Run Downstream biofilm slide was more unexpected. It is difficult to draw conclusions based on a single sample, but it is important to note that the FISH counts showed no viable *C. parvum* or *C. hominis* oocysts, yet the sequencing data

following the removal showed that *C. parvum* oocysts were present. This result could mean that the *C. parvum* oocysts were present on the FISH slide, but were counted as non-viable because they were no longer producing rRNA or because the oocyst excysed. It could also mean that viable oocysts were present on the FISH slide but they were improperly stained by the Cry 1 probe. This result shows how the additional information provided by the FISH to PCR assay can be used to suggest or confirm a risk to human health in a particular water source.

Again, note that the Sandy Run Upstream sample collected on 4/26/11, which was positive by PCR, but negative by FISH, was potentially contaminated. The appearance of amplification product in the negative control which lacked template DNA means that the amplified products of the test DNA must be regarded as suspect and this is why it was omitted from the above analysis. None-the-less, upon sequencing the genotype of *Cryptosporidium* from that sample was determined to be *C. parvum*.

Summary and Conclusions

Sampling by water filtration at Queen Lane WTP discovered a diverse population of *Cryptosporidium* spp. The genotyping showed that there is a potential human health risk associated with using the Schuylkill River as a drinking water source, which currently provides drinking water for over 1.5 million people ("Schuylkill Action Network - Overview", 2008). The sampling by water filtration at Monocacy Creek provided a reference that showed the difference in the detection rates between a water source that is heavily impacted by point sources for *Cryptosporidium* (Queen Lane) and one that is virtually unimpacted (Monocacy Creek).

Based on sequencing data and phylogenetic analysis, the potential risk to human health at the Queen Lane WTP intake is due to the presence of the multiple genotypes of the *C. parvum* species. Of the positive *Cryptosporidium* samples that were sequenced, 75% were species that are associated with human infection. The detection of human infectious genotypes of *Cryptosporidium* at any drinking water source means that there is the potential for a waterborne outbreak of cryptosporidiosis and turbidity standards need to be closely monitored at water treatment plants.

Human infectious genotypes were also found on biofilm slides that were collected at Sandy Run and processed by the FISH to PCR assay. This result shows that the FISH to PCR method can be useful in determining a potential risk to human health in streams that are unimpacted by WWTP discharge and more associated with recreational uses than for their use as a drinking water source. Although the risk is greatly decreased at sources that are not used for drinking water, there is a potential for infection through ingestion while swimming or participating in other water sports.

The diverse population of *Cryptosporidium* genotypes found in this study suggests that multiple sources may be contributing to the parasite population in the Schuylkill River. It is possible that this heterogeneity is attributed to different land uses in different parts of the watershed. Although it is possible for a single oocyst source to release multiple oocyst genotypes, it is more likely that the genetic diversity found at the Queen Lane WTP was the result of multiple sources given the land use profile.

The FISH to PCR assay was established to provide important information regarding the genotyping of oocysts that have already been enumerated by IFA/FISH. The method was tested on environmental samples and it was used to confirm the presence of *C. parvum* at the Queen Lane WTP intake and in Sandy Run. The success of the FISH to PCR assay was not affected by the storage time between FISH and PCR, but in the future, it is advised to run the FISH to PCR assay shortly after the enumeration by IFA/FISH.

The results from the Sandy Run sampling site suggest that the IFA/FISH method alone is not enough to conclude that there is no risk to human health. The FISH to PCR assay provided additional data on the genotypes of the *Cryptosporidium* present in the environmental samples, in addition to the enumeration that was obtained by the IFA/FISH method. Additionally, the FISH to PCR method shows promise in helping to overcome the problems of operator subjectivity and incomplete staining that can occur with IFA/FISH.

Future Work

Because the water filtration method per EPA Method 1622/23 only obtains a small grab sample (<50 L), the sample does not sufficiently reflect the changing conditions of the water source. In the future, the natural tendency of *Cryptosporidium* oocysts to adhere to biofilms

should be utilized in the collection method and the efficiency of the *in situ* biofilm sampler developed by Dr. Wolnyiak should be further investigated. This collection method will allow for a greater understanding of how *Cryptosporidium* populations change over time with respect to changing stream and weather conditions.

Sampling using the biofilm method should take place at unimpacted sites to investigate the possibility of a background concentration of *Cryptosporidium* in eastern Pennsylvania watersheds. Additionally, upstream point sources from the Queen Lane WTP intake should be investigated using the biofilm method. These samples should be collected upstream and downstream from potential point sources to pinpoint the specific sources of human-infectious genotypes of *Cryptosporidium*. Sampling at point sources should also be coupled with fecal sampling.

Finally, the efficacy of the FISH to PCR assay needs to be further investigated, and more precise methods should be used to determine (a) the FISH recovery percentage and (b) the detection limit of the removal. Flow cytometry would greatly enhance the research because it would allow for the sorting of individual oocysts directly onto FISH slides which will allow for the exact detection limit of the removal.

Combining the FISH and PCR methods allows for the more accurate characterization of the *Cryptosporidium* species within a watershed. Accounting for the enhanced collection that is enabled by the biofilm sampling method and the additional molecular data that is generated by the FISH to PCR assay will allow for the more detailed characterization of *Cryptosporidium* populations in source waters in order to locate specific point sources and enact proper control mechanisms.

References

Abelson, H. T; Johnson, L. F.; Penman, S.; Green, H. Changes in RNA in relation to growth of the fibroblast: II. The lifetime of mRNA, rRNA, and tRNA in resting and growing cells. *Cell.* **1974**, 1(4):161-165.

Blewett, D.A. Quantitative techniques in *Cryptosporidium* research. *In*: K.W. Angus and D.A. Blewett (eds.), Proceedings of the 1st International Workshop on Cryptosporidiosis. Edinburgh, Scotland. 1989; p 85-95.

Cacciò, S.M. Molecular epidemiology of human cryptosporidiosis. *Parassitologia*. **2005,** 47(2):185-192.

"CDC - Cryptosporidiosis - Treatment." *Centers for Disease Control and Prevention*. 2 Nov. 2010. Web. 09 Aug. 2011. <http://www.cdc.gov/parasites/crypto/treatment.html>.

Chalmers, R.M. & Giles, M. Zoonotic cryptosporidiosis in the UK - challenges for control. *Journal of Applied Microbiology*. **2010**, 109(5):1487-1497.

Curriero, F. C.; Patz, J. A.; Rose, J. B.; Lele, S. The association between extreme precipitation and waterborne disease outbreaks in the United States, 1948-1994. *American Journal of Public Health.* **2001**, 91:1194-1199.

Current, W.L. & Garcia, L.S. Cryptosporidiosis. *Clinical Microbiology Reviews.* **1991**, 4:325-358.

Di Giovanni, G.D. & LeChevallier, M.W. Quantitative-PCR assessment of Cryptosporidium parvum cell culture infection. *Applied and Environmental Microbiology*. **2005**, 71(3):1495-1500.

Di Giovanni, G. D.; Hoffman, H. M.; Sturbaum, G. D. *Development of a Cryptosporidium Genotyping Method for Regulatory Microscope Slides* (WEF Web Report #4099). Denver, CO: Water Resarch Foundation. 2010. Retrieved November 1, 2010, from http://ntlm.waterrf.org/ProjectsReports/PublicReportLibrary/4099.pdf

Felsenstein, J. Confidence limits on phylogenies: An approach using the bootstrap. *Evolution.* **1985**, 39:783-791.

Feng, Y.; Alderisio, K. A.; Yang, W.; Blancero, L. A.; Kuhne, W. G.; Nadareski, C. A.; Reid, M.; Xiao, L. *Cryptosporidium* genotypes in wildlife from a New York watershed. *Applied and Environmental Microbiology*. **2007**, 73:6475-6483.

Fontaine, M. & Guillot, E. Development of a TaqMan quantitative PCR assay specific for *Cryptosporidium parvum*. FEMS Microbiology Letters, 2002, 214:13–17.

Guy, R.A.; Payment, P.; Krull, U.J.; Horgen, P.A. Real-time PCR for quantification of Giardia and Cryptosporidium in environmental water samples and sewage. *Applied and Environmental Microbiology*. **2003**, 69(9):5178-5185.

Guyot, K.; Follet-Dumoulin, A.; Lelièvre, E.; Sarfati, C.; Rabodonirina, M.; Nevez, G.; Cailliez, J. C.; Camus, D.; Dei-Cas, E. Molecular characterization of *Cryptosporidium* obtained from human in France. *Journal of Clinical Microbiology*. **2001**, 39:3472-3480.

Hall, T.A. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series.* **1999**, 41:95-98.

Hoxie, N.J.; Davis, J.P.; Vergeront, J.M.; Nashold, R.D.; Blair, K.A. Cryptosporidiosis-Associated Mortality Following a Massive Waterborne Outbreak in Milwaukee, Wisconsin. *American Journal of Public Health.* **1997**, 87(12):2032-2035.

Jellison, K.L.; Hemond, H.F.; Schauer, D.B. Sources and Species of *Cryptosporidium* Oocysts in the Wachusett Reservoir Watershed. *Applied and Environmental Microbiology.* **2002**, 68(2):569-575.

Jellison, K.L.; Distel, D.L.; Hemond, H.F.; Schauer, D.B. Phylogenetic Analysis of the Hypervariable Region of the 18S rRNA Gene of *Cryoptosporidium* Oocysts in Feces of Canada Geese (*Branta Canadensis*): Evidence for Five novel Genotypes. *Applied and Environmental Microbiology*. **2004**, 70(1):452-458.

Jellison, K.L.; Lynch, A.E.; Ziemann, J.M. Source Tracking Identifies Deer and Geese as Vectors of Human-Infectious *Cryptosporidium* Genotypes in an Urban/Suburban Watershed. *Environmental Science & Technology.* **2009**, 43:4267-4272.

Johnson, D.W.; Pieniazek, N.J.; Griffin, D.W.; Misener, L.; Rose, J.B. Development of a PCR protocol for sensitive detection of *Cryptosporidium* oocysts in water samples. *Applied and Environmental Microbiology.* **1995**, 61(11):3849-3855.

Kimura, M. A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution.* **1980**, 16:111-120.

Li, H.; Cui, X.; Arnheim, N. Direct electrophoretic detection of the allelic state of single DNA molecules in human sperm by using polymerase chain reaction. *Science.* **1990**, 257:967-971.

Lynch, A. E. 2008. *Cryptosporidium* Source Tracking in the Wissahickon Watershed. M.S. Thesis, Lehigh University.

MacKenzie, W.R.; Hoxie, N.J.; Proctor, M.E.; Gradus, S.M.; Blair, K.A.; Peterson, D.E.; Kazmierczak, J.J.; Addiss, D.G.; Fox, K.R.; Rose, J.B.; Davis, J.P. A Massive Outbreak in Milwaukee of Cryptosporidium Infection Transmitted Through the Public Water Supply. *New England Journal of Medicine.* **1994**, 331(3):161-167.

Mann, H.B. & Whitney, D.R. On a test of whether one of two random variables is stochastically larger than the other. *Annals of Mathematical Statistics.***1947**, 18:50-60.

Mizrachi, L. GenBank: The Nucleotide Sequence Database. In *The NCBI Handbook,* ed. Jo McEntyre and Jim Ostell, 1-15. National Center for Biotechnology Information. 2002. http://www.ncbi.nlm.nih.gov/books/NBK21101/ (accessed January 15, 2011).

Montemayor, M.; Valero, F.; Jofre, J.; and Lucena, F. Occurrence of *Cryptosporidium* spp. oocysts in raw and treated sewage and river water in north-eastern Spain. *Journal of Applied Microbiology*. **2005**, 99:1455–1462.

Morgan, U.M. & Thompson, R.C.A. PCR Detection of Cryptosporidium: The Way Forward?. *Parasitology Today*. **1998**, 14(6):241-245

Pouillot, R.; Beaudeau, P.; Denis, J.B.; Derouin, F.; and AFSSA Cryptosporidium Study Group. A Quantitative Risk Assessment of Waterborne Cryptosporidiosis in France Using Second-Order Monte Carlo Simulation. *Risk Analysis.* **2004**, 24 (1):1-17.

Ramos, F.; Zurabian, R.; Moran, P.; Ramiro, M.; Gomez, A.; Clark, C.G.; Melendro, E.I.; Garcia, G.; Ximenez, C. The effect of formalin fixation on the polymerase chain reaction characterization of Entamoeba histolytica. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **1999**, 93(3):335-336.

Robertson, L.J.; Campbell, A.T; Smith, H.V. Survival of Cryptosporidium parvum oocysts under various environmental pressures. *Applied and Environ Microbiology*. **1992**, 58:3494-3500.

Scallan, E.; Hoekstra, R.M.; Angulo, F.J.; Tauxe, R.V.; Widdowson, M.A.; Roy, S.L.; Jones, J.L.; Griffin, P.M. Foodborne illness acquired in the United States--major pathogens. *Emerging Infectious Diseases*. **2011**, 17(1):7-15.

"Schuylkill Action Network - Overview." *Schuylkill Action Network - Home*. 2008. Web. 10 Aug. 2011. <http://www.schuylkillwaters.org/schuylkill.cfm>.

Shapiro, S.S. & Wilk, M.B. An analysis of variance test for normality (complete samples). *Biometrika.* **1965,** 52(3-4):591-611.

Shi, K.; Jian, F.; Lv, C.; Ning, C.; Zhang, L.; Ren, X.; Dearen, T.K.; Li, N.; Qi, M.; Xiao, L. Prevalence, genetic characteristics, and zoonotic potential of Cryptosporidium species causing infections in farm rabbits in China. *Journal of Clinical Microbiology*. **2010**, 48:3263–3266.

Tamura, K.; Peterson, D.; Peterson, N.; Stecher, G.; Nei, M.; Kumar, S. MEGA5: Molecular Evolutionary Genetics Analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution*. **2011**, first published online May 4, 2011. doi:10.1093/molbev/msr121

Thompson, C.A.; Olson, M.E.; Zhu, G.; Enomoto, S.; Abrahamsenm, M.S.; Hijjawi, N.S. Cryptosporidium and Cryptosporidiosis. *Advances in Parasitology*. **2005**, 59:77-158.

USEPA. 2006. Source Water Monitoring Guidance Manual for Public Water Systems for the Final Long Term 2 Enhanced Surface Water Treatment Rule. EPA 815-R06-005, U.S. Environmental Protection Agency, Office of Water, Washington, DC.

USEPA. 2005. Method 1622: Cryptosporidium in Water by Filtration/IMS/FA. EPA 815-R-05-001, U.S. Environmental Protection Agency, Office of Water, Washington, DC.

Vesey, G.; Ashbolt, N.; Fricker, E.J.; Deere, D.; Williams, K.L.; Veal, D.A.; Dorsch, M. The use of a ribosomal RNA targeted oligonucleotide probe for fluorescent labelling of viable *Cryptosporidium* parvum oocysts. *Journal of Applied Microbiology*. **1998**, 85(3):429-440.

Wolyniak DiCesare, E. 2011. Impact of biofilm and other environmental factors on the fate and transport of *Cryptosporidium* oocysts. Ph.D. dissertation, Lehigh University.

WHO. Risk Assessment of *Cryptosopridium* in Drinking Water. World Health Organization, Geneva, Switzerland. 2009.

Xiao, L.; Escalante, L.; Yang, C.; Sulaiman, I.; Escalante, A.A.; Montali, R.J.; Fayer, R.; Lal, A.A. Phylogenetic analysis of Cryptosporidium parasites based on the small subunit ribosomal RNA gene locus. *Applied Environmental Microbiology*. **1999**, 65:1578–1583

Xiao, L.; Fayer, R.; Ryan, U.; Upton, S.J.. *Cryptosporidium* Taxonomy: Recent Advances and Implications for Public Health. *Clinical Microbiology Reviews*. **2004**, 17:72-97.

Xiao, L.; Bern, C.; Sulaiman, I.M.; Lal, A.A. Molecular epidemiology of human cryptosporidiosis. In: Thompson, R.C.A.; Armson, A.; Ryan, U.M. (eds.). *Cryptosporidium: from molecules to disease*. Elsevier, Amsterdam, The Netherlands, 2004, pp. 121-146.

Zhou, A.; Singh, A.; Jiang, J.; Xiao, L. Molecular surveillance of *Cryptosporidium* spp. in raw waste water in Milwaukee: implications for understanding outbreak occurrence and transmission dynamics. *Journal of Clinical Microbiology*. **2003**, 41:5254-5257

Ziemann, J. M. 2006. DNA source tracking of *Cryptosporidium* spp. oocysts in the Wissahickon Creek Watershed of Philadelphia, PA. M.S. thesis, Lehigh University.

Appendix A: Sampling locations

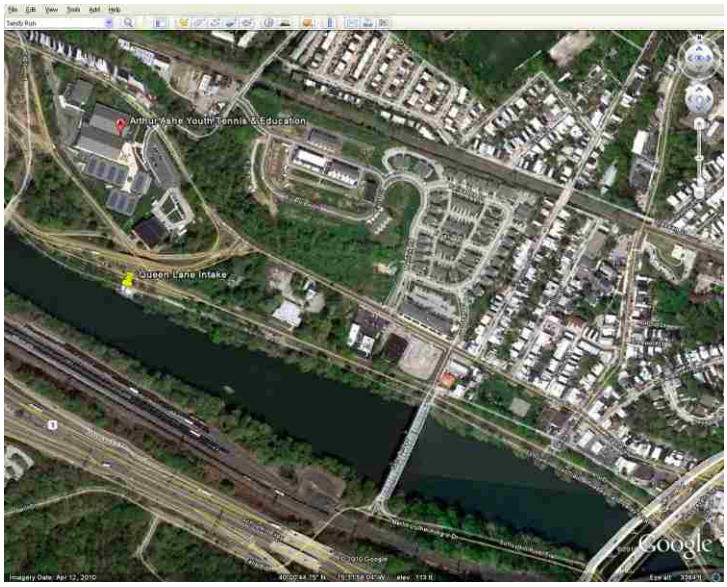

Figure A-1. *Queen Lane Intake* (courtesy of Google Earth). The Queen Lane intake sampling site was located off of Kelly Drive, near the East Falls neighborhood of Philadelphia, PA. The closest address for reference was the Arthur Ashe Youth Tennis & Education center (labeled as point A on the map).

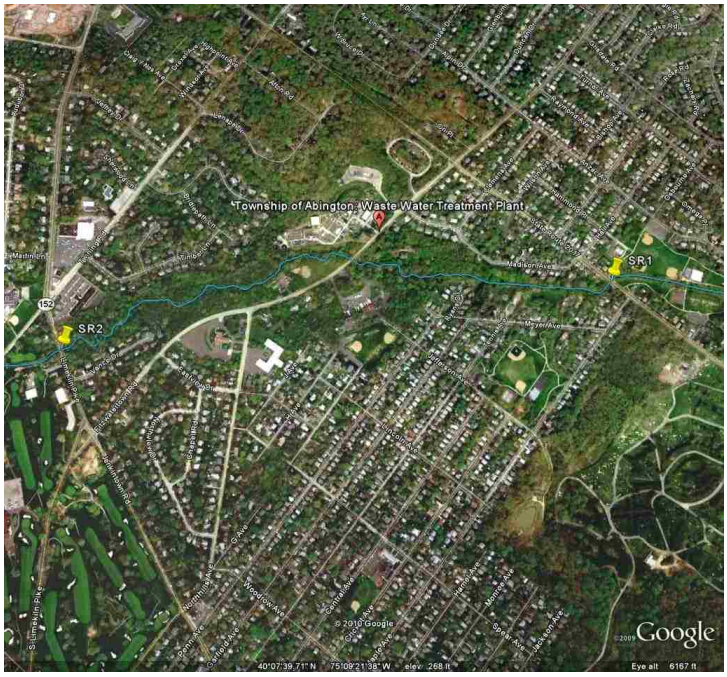

Figure A-2. *Sandy Run - Upstream and Downstream from the Township of Abington Wastewater Treatment Plant* (courtesy of Google Earth). The Abington WWTP is shown as point A.

Figure A-3. *SR1 - Sandy Run Upstream* (courtesy of Google Earth). SR1 was located on the upstream side of Abington WWTP, where Sandy Run intersects State Route 2017/Susquehanna Road. There was a stormwater outfall located a few yards downstream of where the biofilms were installed. The biofilm sampler was tied to a secure tree root at this site.

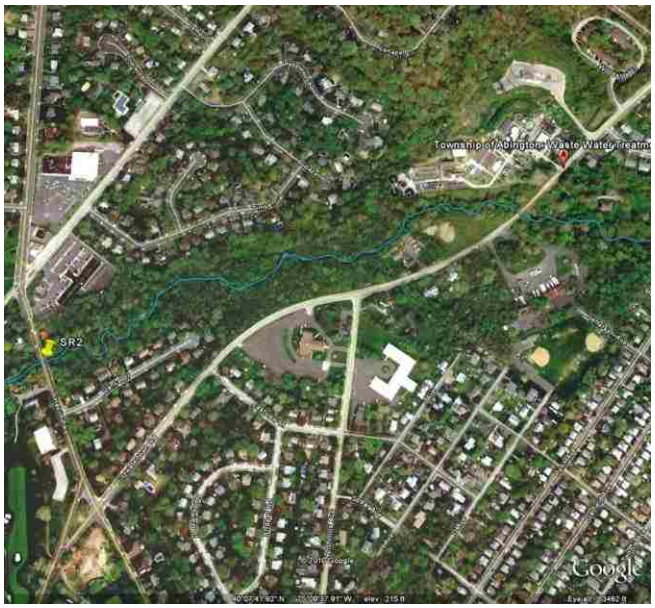

Figure A-4. *SR2 - Sandy Run Downstream* (courtesy of Google Earth). SR2 was located downstream of Abington WWTP, where Sandy Run intersects Limekiln Pike. The biofilm sampler was located on the upstream side of a bridge that runs across Sandy Run. The sampler was secured using a stake that is hammered into the streambed.

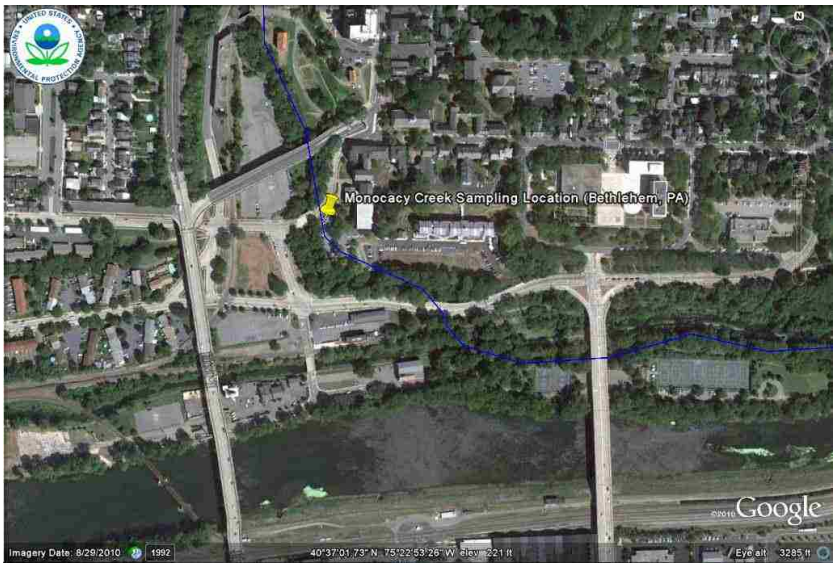

Figure A-5 *Monocacy Creek - Bethlehem, PA* (courtesy of Google Earth). The Monocacy Creek Downstream sampling site was located under the Spring St. Bridge in Historic Bethlehem.

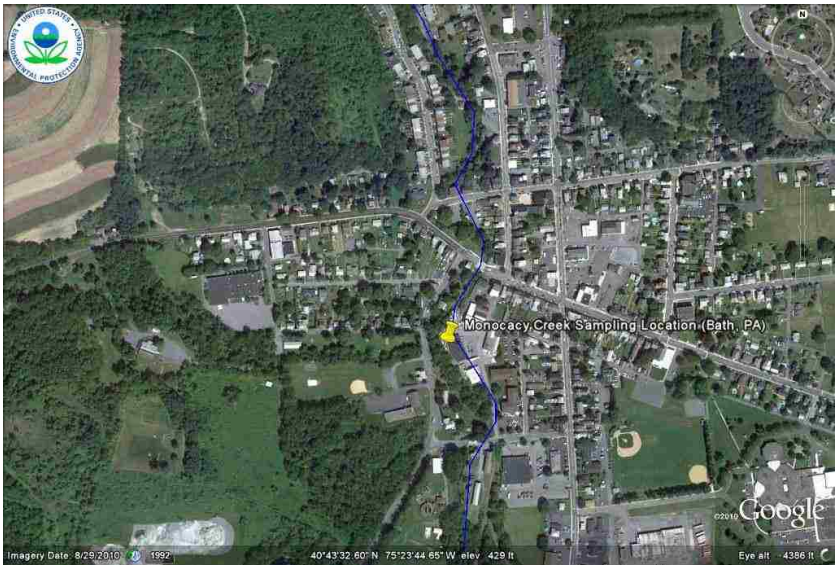

Figure A-6 *Monocacy Creek - Bath, PA* (courtesy of Google Earth). The Monocacy Creek Upstream sampling site was located in Bath, PA off of Race St (PA route 987/329). It was across the street from Keystone Park. The biofilm holder was anchored to a protruding tree route.

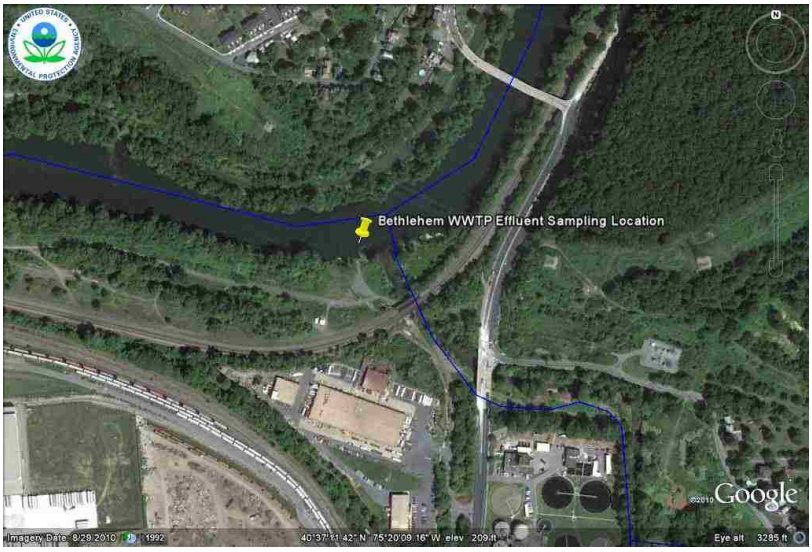

Figure A-7 *Bethlehem Wastewater Treatment Plant Effluent Sampling Sites - Lehigh River* (courtesy of Google Earth). The Bethlehem WWTP sampling site was located in the Lehigh River just upstream from its confluence with Saucon Creek. The point shown on the map is the location of the discharge line. Biofilm Samplers were attached to old brake rotors so that they would not be carried downstream. They were submerged approximately 100 ft upstream from the WWTP effluent discharge and approximately 50 ft downstream from the WWTP effluent discharge.

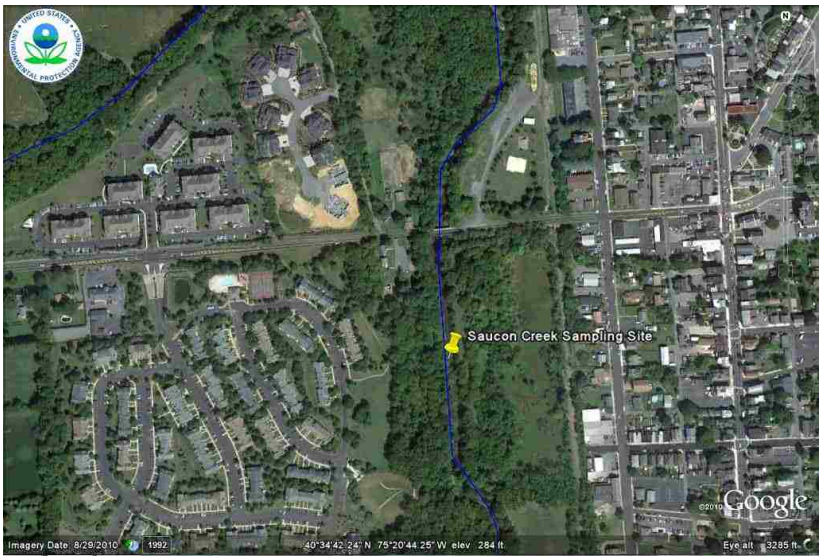

Figure A-8 *Saucon Creek* (courtesy of Google Earth). The Saucon Creek sampling site was located approximately 600 ft upstream from the Friedensville Road/Water St Bridge. This location borders Saucon Valley, PA and Hellertown, PA. The biofilm sampler was anchored to a tree.

Appendix B: PCR cocktail and primer information

Table B-1. Reagents for PCR cocktail for one sample

Initial PCR	Secondary PCR
1.85 µL H_2O	30.85 µL H_2O
8.75 µL $MgCl_2$	8.75 µL $MgCl_2$
5 µL Buffer	5 µL Buffer
0.75 µL DNTP (x4)	0.75 µL DNTP (x4)
0.5 µL KLJ1 & 0.5 µL KLJ2	0.5 µL CPB-DIAGF & 0.5 µL CPB-DIAGR
0.4 µL Taq	0.4 µL Taq

Table B-2. Primers used for nested PCR

KLJ1	5'-CCACATCTAAGGAAGGCAGC-3'
KLJ2	5'-ATGGATGCATCAGTGTAGCG-3'
CPB-DIAGF	5'-AAGCTCGTAGTTGGATTTCTG-3'
CPB-DIAGR	5'-TAAGGTGCTGAAGGAGTAAGG-3'

Appendix C: PCR results visualized by gel electrophoresis

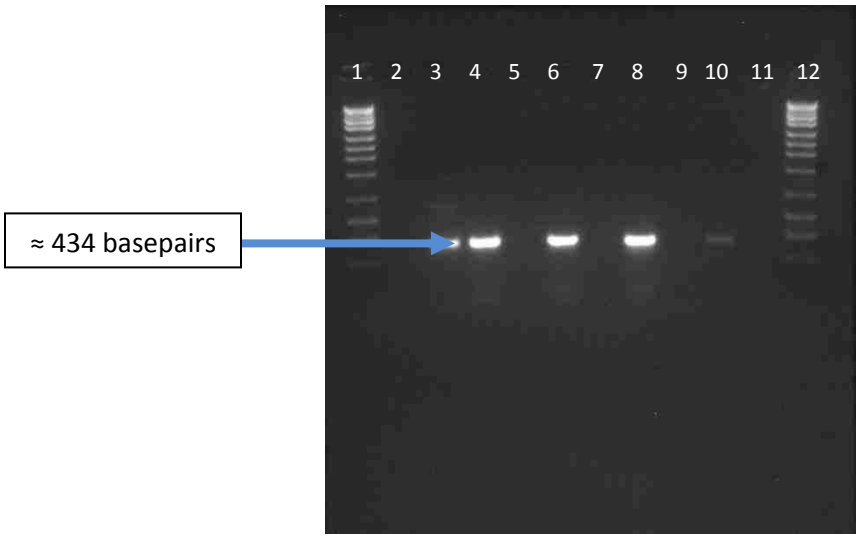

Figure C-1. UV photograph of 2/23/11 Queen Lane water filter sample after 75 minutes of gel electrophoresis in a 1.4% agarose gel. From left to right, the lanes are as follows (1) 1 kilobase pair ladder, (2) 2/23/11 QLA Sample, (3) 2/23/11 QLB Sample, (4) IMS positive control, (5) IMS negative control, (6) DNA extraction positive control, (7) DNA extraction negative control, (8) primary PCR positive control, (9) primary PCR negative control, (10) secondary PCR positive control, (11) secondary PCR negative control, and (12) an additional 1kbp ladder. Notice that the 2/23/11 QLB Sample has a bright band at approximately 434 basepairs and a faint band at a higher position in the lane.

Figure C-2. UV photograph of 4/26/11 Queen Lane water filter samples after 60 minutes of gel electrophoresis in a 1.4% agarose gel. From left to right, he lanes are as follows (1) 1 kilobase pair ladder, (2) 4/26/11 QLA Sample, (3) 4/26/11 QLA Sample, (4) 4/26/11 QLB Sample, (5) IMS positive control, (6) IMS negative control, (7) DNA extraction positive control, (8) DNA extraction negative control, (9) primary PCR positive control, and (10) an additional 1kbp ladder (primary PCR negative control, secondary PCR positive control, and secondary PCR negative control are not shown). The QLA sample was processed by DNA extraction and nested PCR in duplicate,

hence the two wells representing the 4/26/11 sample. Notice that although both lane 3 and lane 4 appear to have a band at the same location as the positive controls, sequencing proved that *Cryptosporidium* was not present. Instead, the primers misread an endosymbiotic diatom.

Appendix D: Staining in solution protocol for FISH

1. Started with 100 µL of oocyst suspension after hybridization of the Cry 1 probe
2. Added one drop of Merifluor detection reagent and one drop of counterstain to each sample and vortexed
3. Incubated in the dark for 30 minutes at 4 ^{o}C
4. Centrifuged for 4 min at 13,000 rpm and discarded the supernatant
5. Added 500 µL of Millipore water and vortexed to rinse oocysts
6. Centrifuged for 4 min at 13,000 rpm and discarded the supernatant
7. Resuspended in 20 µL of Millipore water and transferred to Merifluor slide
8. Dried slides in the dark at room temperature
9. Fixed slides with 10 µL 100% methanol and allowed to dry at room temperature
10. Added one drop of formalin-free mounting media between each well and applied cover slip

Appendix E: Cloning and restriction enzyme digestion results visualized by gel electrophoresis

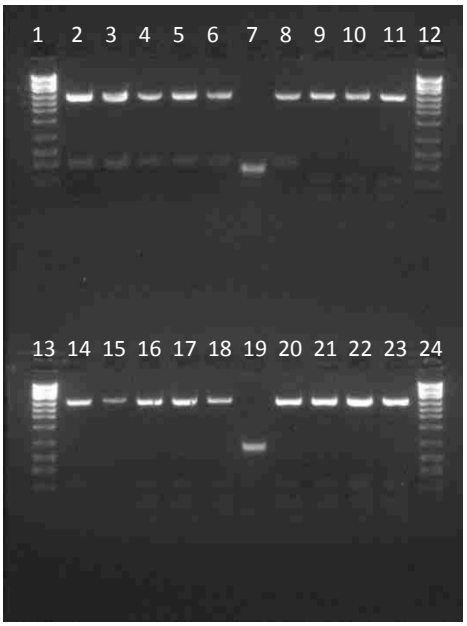

Figure E-1. UV photograph taken on 1/7/11 of clones of the 9/28/10 QLB and 11/23/10 QLB samples after 60 minutes of gel electrophoresis in a 1.4% agarose gel. The positive controls are in lanes 7 and 19. Note that lanes 2-6 and 8 did not cut with *Nde*I and they are slightly higher than the control. Sequencing confirmed that they were not *Cryptosporidium*. Lanes 9-11, 14-18 and 20-23 all cut with *Nde*I. Sequencing determined that they were *C. parvum.*

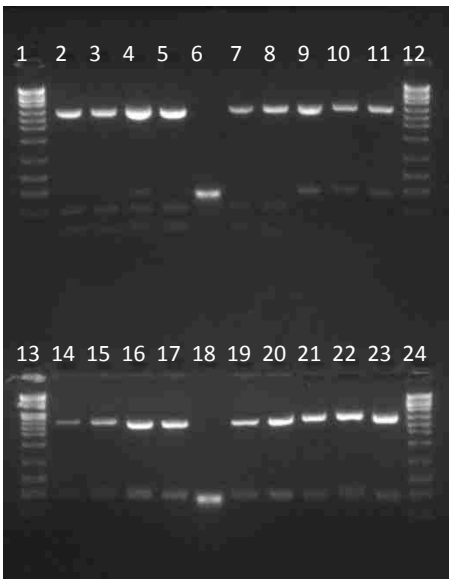

Figure E-2. UV photograph taken on 2/21/11 of clones of the 1/11/11 QLB and 1/25/11 QLB samples after 60 minutes of gel electrophoresis in a 1.4% agarose gel. The positive controls are in lanes 6 and 18. Note that lanes 2, 3, 5, 7 and 8 all cut with *Nde*I. Sequencing determined that

these were closest related to *C. suis*. Lanes 9-11, 14-17 and 19-23 did not cut with *Nde*I and, although they appear to be higher than the controls, sequencing confirmed that they were the goose I genotype of *Cryptosporidium.* Also note the incomplete digestion of lane 4. This clone was sequenced and it turned out to be the exact same sequence as the other clones from the 1/11/11 QLB sample.

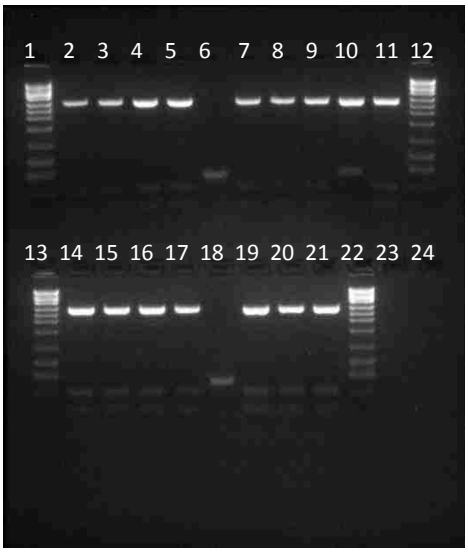

Figure E-3. UV photograph taken on 7/26/11 of clones from the 3/24/11 SR2 and 3/10/11 QL FISH slides after 60 minutes of gel electrophoresis in a 1.4% agarose gel. Positive controls are in lanes 6 and 18. Lane 10 is the only lane not to have cut with *Nde*I and this is most likely the result of incomplete digestion since all of the clones that were sequenced were confirmed to be *C. parvum.*

Appendix F: Sequence alignments

Table F-1. Sequence alignment for all consensus clones from positive samples.

	10 20 30 40 50 60 70 80 90 100
	....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
9/28/10 QLA All Clones	AAGCTCGTAGTTGGATTTCTGTTAATAATTTATATAARATATTTTGATGAATATTTATATAATATTAACATAATTCATATTACTATATATTTTAGTATAT
11/23/10 QL All Clones	AAGCTCGTAGTTGGATTTCTGTTAATARTTTATATAAAATATTTTGATGAATATTTATATAATATYAACATAATTCATATTACTATATATTTTAGTATAT
1/11/11 QLA All Cones	AAGCTCGTAGTTGGATTTCTGTTAATAATTTATATAAAATATTTTGAATATTTATATAACATTAACATAATTCATATTACTATATTTTTTAGTATATGAA
1/11/11 QLB All Clones	AAGCTCGTAGTTGGATTTCTGTTAATAATTTAYATATAATATTTTTAATATTTATATAATATTAACATAATTCATATTACTATATTATTATTAGTATATG
1/25/11 QL Clone 2	AAGCTCGTAGTTGGATTTCTGCTAATTTTTGCATACGATACCACGGTATTTATGTAAAATTAGCATAATCCGCATTACCTCGCGTATGCGGAATTTTACT
1/25/11 QL Clones 1&3	AAGCTCGTAGTTGGATTTCTGCTAATTTTTGCATACAATACCACGGTATTTATGTAAAATTAGCATAATCCGCATTACCTCGCGTATGCGGAAYTTTACT
2/23/11 QL Clone 1	AAGCTCGTAGTTGGATTTCTGTTGTATAATTTTATAATATTACCAAGGTAATTATTATATTATCAACATCCTTCCTATTATATTCTAAATATATAGGAAA
2/23/11 QL Clone 2	AAGCTCGTAGTTGGATTTCTGTTGTATAATTTATAATATTACCAAGGTAATTATTATATTATCAACATCCTTCCTATTATATTCTAAATATATAGGAAAT
3/8/11 QL Biofilm Clone 1	AAGCTCGTAGTTGGATTTCTGTTAATAATTTATATAAAATATTTTGATGAATATTTATATAATATTAACATAATTCATATTACTACATATTTTAGTATAT
3/8/11 QL Biofilm All	AAGCTCGTAGTTGGATTTCTGTTAATAATTTATATAAAATATTTTGATGAATATTTRTATAATATTAACATWATTCATATTACTAYATATTTTAGTATAT
3/22/11 SR2 Biofilm All	AAGCTCGTAGTTGGATTTCTGTTAATAATTTATATAAAATATTTTGATGAATATTTATATAATATTAACATAATTCATATTACTATATATTTTAGTATAT
4/26/11 SR1 Biofilm All	AAGCTCGTAGTTGGATTTCTGTTAATAATTTATATAAAATATTTTGATGAATATTTATATAATATTAACATAATTCATATTACTATATATTTTAGTATAT
4/26/11 QL Biofilm Clone 1	AAGCTCGTAGTTGGATTTCTGTTAATAATTTATATAAAATATTTTGATGAATATCTATATAATATTAACATAATTCATATTACTATATATTTTAGTATAT
4/26/11 QL Biofilm Clone2&4	AAGCTCGTAGTTGGATTTCTGTTAATAATTTATATAAAATATTTTGATGAATATTTATATAATATTAACATAATTCATATTACTATATATTTTAGTATAT
4/26/11 QL Biofilm Clone 3	AAGCTCGTAGTTGGATTTCTGTTAATAATTTATATAAAATATTTTGATGAATATTTATATAATATTAACATAATTCATATTACTATATATTTTAGTATAT

	110 120 130 140 150 160 170 180 190 200 ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
9/28/10 QLA All Clones	GAAATTTTACTTTGAGAARATTAGAGTGCTTAAAGCAGGCATATGCCTTGAATACTCCAGCATGGAATAATATTAAAGATTTTTATCTTTCTTATTGGTT
11/23/10 QL All Clones	GAAATTTTACTTTGAGAAAATTAGAGTGCTTAAAGCAGGCATATGCCTTGAATACTCCAGCATGGAATAATATTAAAGATTTTTATCTTTCTTATTGGTT
1/11/11 QLA All Cones	ATTTTACTTTGAGAAAATTAGAGTGCTTAAAGCAGGCATATGCCTTGAATACTCCAGCATGGAATAATATTAAAGATTTTTATCTTTCTTATTGGTTCTA
1/11/11 QLB All Clones	AAATTTTACTTTGAGAAAATTAGAGTGCTTAAAGCAGGCATATGCCTTGAATACTCCAGCATGGAATAATATAAAAGATTTTTATCTTTTTTATTGGTTC
1/25/11 QL Clone 2	TTGAGAAAATTAGAGTGCTTAAAGCAGGCAATTGCCTTGAATACTCCAGCATGGAATAATATTAAGGATTTTTATCCTTTTTATTGGTTCTAGGATAAAA
1/25/11 QL Clones 1&3	TTGAGAAAATTAGAGTGCTTAAAGCAGGCAATTGCCTTGRATACTCCAGCATGGARTAATATTAAGGATTTTTATCCTTTTTATTGGTTCTAGGATAAAA
2/23/11 QL Clone 1	TTTTACTTTGAGAAAATTAGAGTGCTTAAAGCAGGCAACTGCCTTGAATACTCCAGCATGGAATAATAAGTAAGGACTTTTGTCTTTCTTATTGGTTCTA
2/23/11 QL Clone 2	TTTACTTTGAGAAAATTAGAGTGCTTAAAGCAGGCAACTGCCTTGAATACTCCAGCATGGAATAATAAGTAAGGACTTTTGTCTTTCTTATTGGTTCTAG
3/8/11 QL Biofilm Clone 1	GAAATTTTACTTTGAGAAAATTAGAGTGCTTAAAGCAGGCATATACCTTGAATACTCCAGCATGGAATAATATTAAAGATTTTTATCTTTCTTATTGGTT
3/8/11 QL Biofilm All	GAAATTTTACTTTGAGAAAATTAGAGTGCTTAAAGCAGGCATATRCCTTGAATACTCCAGCATGGWATAATATTAAAGATTTTTATYTTTCTTATTGGTT
3/22/11 SR2 Biofilm All	GAAATTTTACTTTGAGAAAATTAGAGTGCTTAAAGCAGGCATATGCCTTGAATACTCCAGCATGGAATARTATTAAAGATTTTTATCTTTCTTATTGGTT
4/26/11 SR1 Biofilm All	GAAATTTTACTTTGAGAAAATTAGRGTGCTTAAAGCAGGCATATGCCTTGAATACTCCAGCATGGARTAATATTAAAGATTTTTATCTTTCTTATTGGTT
4/26/11 QL Biofilm Clone 1	GAAATTTTACTTTGAGAAAATTAGAGTGCTTAAAGCAGGCATATGCCTTGAATACTCCAGCATGGAATAGTATTAAAGATTTTTATCTTTCTTATTGGTT
4/26/11 QL Biofilm Clone2&4	GAAATTTTACTTTGAGAAAATTAGAGTGCTTAAAGCAGGCATRTGCCTTGAATACTCCAGCATGGAATAATATTAAAGATTTTTATCTTTCTTATTGGTT
4/26/11 QL Biofilm Clone 3	GAAATTTTACTTTGAGAAAATTAGAGTGCTTAAAGCAGGCATATGCCTTGAATACTCCAGCAAGGAATAATATTAAAGATTTTTATCTTTCTTATTGGTT

	210 220 230 240 250 260 270 280 290 300
	....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
9/28/10 QLA All Clones	CTAAGATAAGAATAATGATTAATAGGGACAGTTGGGGGCATTTGTATTTAACAGTCAGAGGTGAAATTCTTAGWTTTGTTAAAGACAAACTAATGCGAAA
11/23/10 QL All Clones	CTAAGATAAGAAYRATGATTAATAGGGACAGTTGGGGGCATTTGTATTTAACAGTCAGAGGTGAAATTCTTAGATTTGTTAAAGACAAACTAATGCGAAA
1/11/11 QLA All Cones	AGATAAGAATAATGATTAATAGGGACAGTTGGGGGCATTTGTATTTAACAGTCAGAGGTGAAATTCTTAGATTTGTTAAAGACAAACTAATGCGAAAGCA
1/11/11 QLB All Clones	TAAGATAAAAATAATGATTAATAGGGACAGTTGGGGGCATTTGTATTTAACAGTCAGAGGTGAAATTCTTAGATTTGTTAAAGACAAACTAGTGCGAAAG
1/25/11 QL Clone 2	ATAATGATTAATAGGGACAGTTGGGGGCATTTGTATTTAACAGTCAGAGGTGAAATTCTTAGATTTGTTAAAGACAAACTACTGCGAAAGCATTTGCCAA
1/25/11 QL Clones 1&3	ATAATGATTAATAGGGACAGTTGGGGGCATTTGTATTTAACAGTCAGAGGTGAAATTCTTAGATTTGTTAAAGACAAACTACTGCGAAAGCATTTGCCAA
2/23/11 QL Clone 1	GGACAAAAGTAATGGTTAATAGGGACAGTTGGGGGCATTCGTATTTAACAGCCAGAGGTGAAATTCTTAGATTTGTCAAAGACGAACTACTGCGAAAGCA
2/23/11 QL Clone 2	GACAAAAGTAATGGTTAATAGGGACAGTTGGGGGCATTCGTATTTAACAGCCAGAGGTGAAATTCTTAGATTTGTTAAAGACGAACTACTGCGAAAGCAT
3/8/11 QL Biofilm Clone 1	CTAAGATAAGAATAATGATTAATAGGGACAGTTGGGGGCATTTGTATTTAACAGTCAGAGGTGAAATTCTTAGATTTGTTAAAAACAAACTAATGCGAAA
3/8/11 QL Biofilm All	CTAAGATAAGAATAATGATTAATAGGGACAGTTGGGGGCATTTGTATTYAACAGTCAGAGGTGAAATTCTTAGATTTGTTAAARACAAACTAATGCGAAA
3/22/11 SR2 Biofilm All	CTAAGATAAGAATAATGATTAATAGGGACAGTTGGGGGCATTTGTATTTAACAGTCAGAGGTGAAATTCTTAGATTTGTTAAAGACAAACTAATGCGAAA
4/26/11 SR1 Biofilm All	CTAAGATAAGAATAATGATTAATAGGGACAGTTGGGGGCATTTGTATTTAACAGTCAKAGGTGAAATTCTTAGATTTGTTAAAGACAAACTAATGCGAAA
4/26/11 QL Biofilm Clone 1	CTAAGATAAGAATAATGATTAATAGGGACAGTTGGGGGCATTTGTATTTAACAGTCAGAGGTGAAATTCTTAGATTAGTTAAAGACAAACTAATGCGAAA
4/26/11 QL Biofilm Clone2&4	CTAAGATAAGAATAATGATTAATAGGGACAGTTGGGGGCATTTGTATTTAACAGTCAGAGGTGAAATTCTTAGATTTGTTAAAGACAAACTAATGCGAAA
4/26/11 QL Biofilm Clone 3	CTAAGATAAGGATAATGATTAATAGGGACAGTTGGGGGCATTTGTATTTAACAGTCAGAGGTGAAATTCTTAGATTTGTTAAAGACAAACTAATACGAAA

	310 320 330 340 350 360 370 380 390 400
	....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
9/28/10 QLA All Clones	GCATTTGCCAAGGATGTTTTCATTAATCAAGAACGAAAGTTAGGGGATCGAAGACGATCAGATACCGTCGTAGTCTTAACCATAAACTATGCCAACTAGA
11/23/10 QL All Clones	GCATTTGCCAAGGATGTTTTCATTAATCAAGAACGAAAGTTAGGGGATCGAAGACGATCAGATACCGTCGTAGTCTTAACCATAAACTATGCCAACTAGA
1/11/11 QLA All Cones	TTTGCCAAGGATGYTTTCATTAATCAAGAACGAAAGTTAGGGGATCGAAGACGATCAGATACCGTCGTAGTCTTAACCATAAACTATGCCAACTAGAGAT
1/11/11 QLB All Clones	CATTTGCCAAGGATGTTTTCATTAATCAAGAACGAAAGTTAGGGGATCGAAGACGATCAGATACCGTCGTAGTCTYAACCATAAACTATGCCAACTAGAG
1/25/11 QL Clone 2	GGATGTTCTCATTAATCAAGAACGAAAGTTAGGGGATCGAAGACGATCAGATACCGTCGTAGTTTTAACCATAAACTATGCCGACTAGAGATTGGAGGTT
1/25/11 QL Clones 1&3	GGATGTTTTCATTAATCAAGAACGAAAGTTAGGGGATCGAAGACGATCAGATACCGTCGTAGTCTTAACCATAAACTATGCCGACTAGAGATTGGAGGTT
2/23/11 QL Clone 1	TTTGCCAAGGATGTTTTCATTAATCAAGAACGAAAGTTAGGGGATCGAAGACGATCAGATACCGTCGTAGTCTTAACCATAAACTATGCCGACTAGAGAT
2/23/11 QL Clone 2	TTGCCAAGGATGTTTTCATTAATCAAGAACGAAAGTTAGGGGATCGAAGACGAACAGATACCGTCGTAGTCTTAACCATAAACTATGCCGACTAGAGATT
3/8/11 QL Biofilm Clone 1	GCATTTGCCAAGGATGTTTTCATTAATCAAGAACGAAAGTTAGGGGATCGAAGACGATCAGATACCGTCGTAGTCTTAACCATAAACTATGCCAACTAGA
3/8/11 QL Biofilm All	GCATTTGCCAAGGATGTTTTCATTAATCAAGAACGAAAGTTAGGGGATCGAAGACGATCAGATACCGTCGTAGTCTTAACCATAAACTATGCCAACTAGA
3/22/11 SR2 Biofilm All	GCATTTGCCAAGGATGTTTTCATTAATCAAGAACGAAAGTTAGGGGATCGAAGACGATCAGATACCGTCGTARTCTTAACCATAAACTATGCCAACTAGA
4/26/11 SR1 Biofilm All	GCATTTGCCAAGGATGTTTTCATTAATCAAGAACGAAAGTTAGGGGATCGAAGACGATCAGATACCGTCGTAGTCTTAACCAWAAACTATGCCAACTAGA
4/26/11 QL Biofilm Clone 1	GCATTTGCCAAGGATGTTTTCATTAATCAAGAACGAAAGTTAGGGGATCGAAGACGATCAGATACCGTCGTAGTCTTAACCATAAACTATGCCAACTAGA
4/26/11 QL Biofilm Clone2&4	GCATTTGCCAAGGATGTTTTCATTAATCAAGAACGAAAGTTAGGGGATCGAAGACGATCAGATACCGTCGTAGTCTTAACCATAAACTATGCCAACTAGA
4/26/11 QL Biofilm Clone 3	GCATTTGCCAAGGATGTTTTCATTAATCAAGAACGAAAGTTAGGGGATCGAAGACGATCAGATACCGTCGTAGTCTTAACCATAAACTATGCCAACTAGA

	410 420 430 ....|....|....|....|....|....|....|
9/28/10 QLA All Clones	GATTGGAGGTTGTTCCTTACTCCTTCAGCACCTTA
11/23/10 QL All Clones	GATTGGAGGTTGTTCCTTACTCCTTCAGCACCTTA
1/11/11 QLA All Cones	TGGAGGTTGTTCCTTACTCCTTCAGCACCTTA
1/11/11 QLB All Clones	ATTGGAGGTTGTTCCTTACTCCTTCAGCACCTTA
1/25/11 QL Clone 2	GTTCCTTACTCCTTCAGCACCTTA
1/25/11 QL Clones 1&3	GTTCCTTACTCCTTCAGCACCTTA
2/23/11 QL Clone 1	TGGAGGTTGTTCCTTACTCCTTCAGCACCTTA
2/23/11 QL Clone 2	GGAGGTTGTTCCTTACTCCTTCAGCACCTTA
3/8/11 QL Biofilm Clone 1	GATTGGAGGTTGTTCCTTACTCCTTCAGCACCTTA
3/8/11 QL Biofilm All	GATTGGAGGTTGTTCCTTACTCCTTCAGCACCTTA
3/22/11 SR2 Biofilm All	GATTGGAGGTTGTTCCTTACTCCTTCAGCACCTTA
4/26/11 SR1 Biofilm All	GATTGGAGGTTGTTCCTTACTCCTTCAGCACCTTA
4/26/11 QL Biofilm Clone 1	GATTGGAGGTCGTTCCTTACTCCTTCAGCACCTTA
4/26/11 QL Biofilm Clone2&4	GATTGGRGGTTGTTCCTTACTCCTTCAGCACCTTA
4/26/11 QL Biofilm Clone 3	GATTAGAGGTTGTTCCTTACTCCTTCAGCACCTTA

Appendix G: Statistical methods

Shapiro-Wilk test

$$W = \left\{\frac{b}{S\sqrt{n-1}}\right\}^2$$

$$b = \sum_{i=1}^{k} a_{n-i+1}\left(x_{n-i+1} - x_{(i)}\right) = \sum_{i=1}^{k} b_i$$

Where $x_{(i)}$ represents the smallest ordered value in the sample, and a_i is the "coefficient for Shapiro-Wilk *W*-test for normality" which is a calculated value that is dependent on the sample size n and is found in a table computed by Shapiro and Wilk. k is the greatest integer less than or equal to $n/2$.

The normality of the data should be rejected if W is lower than the critical value for the sample size n, which is also found in a table computed by Shapiro and Wilk.

Variance ratio test based on the F-distribution

$$F\ ratio = \frac{\max[S_1^2, S_2^2]}{\min[S_1^2, S_2^2]}$$

Where S_1^2 and S_2^2 represent the sample variances for two populations of interest.

The *F* ratio is compared to the *F* distribution based on the degrees of freedom and values of α. If *F* ratio > *F* distribution, the student *t*-test for similarity between two populations will not hold.

Significance testing using the student's t test comparing two samples with similar variances

$$t^* = \frac{|\overline{X_1} - \overline{X_2}|}{S_m}$$

$$S_m = \sqrt{\hat{S}^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

$$\hat{S}^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

Where n_1 and n_2 represent the number of samples per population, S_1^2 and S_2^2 represent the sample variances for two populations, $\hat{S}^2$ is the pooled variance, and S_m is the standard error of the mean. The degree of freedom is calculated as: $df = n_1 + n_2 - 2$. After calculating the test statistic, t^*, compare it to the critical value and if $t^* > t^c_{\alpha,df}$ you reject the null hypothesis.

Mann-Whitney Test[1]

The Mann-Whitney test is a nonparametric procedure used to compare the median of two samples, say sample *X* and sample *Y*. To compute the test statistic, the two samples must be combined and the observations must be ranked from smallest to largest while keeping track

[1] Daniel, Wayne W. "Nonparametric and Distribution-free Statistics." *Biostatistics: a Foundation for Analysis in the Health Sciences*. Hoboken, NJ: Wiley, 2005. 680-762. Print.

of the sample to which each observation belongs. Tied observations are assigned a rank equal to the mean of the rank positions for which they are tied. The test statistic is:

$$T = S - \frac{n(n+1)}{2}$$

Where n is the number of sample *X* observations and *S* is the sum of the ranks assigned to the samples observations from the population of *X* values. If the hypothesis is that the median of *X* is larger than the median of *Y*, the sum of the ranks assigned to the observations from the *X* population should be smaller than the sum of the ranks assigned to the observations from the *Y* population. The test statistic as computed by Mann and Whitney is based on this rationale so that a sufficiently small value of *T* will cause rejection of the hypothesis that the median of *X* is larger than the median of *Y*.

Vita

Contact Information

Colin McLeod
203-219-9888
cmm504@gmail.com

Personal Information

Date of Birth: December 2, 1986
Place of Birth: Stamford, Connecticut
Parents: Chris and Elaine McLeod

Employment History

Lehigh University, Department of Civil and Environmental Engineering, Bethlehem, PA
2010 - 2011
Graduate Student Researcher
Conducted independent research under the auspices of the water department of a large city in order to identify and study *Cryptosporidium* genotypes in eastern Pennsylvania watersheds.

United States Environmental Protection Agency, Region V Air Enforcement and Compliance Assurance Branch, Chicago, IL
2009-2010
Environmental Engineer Intern
Provided technical and non-technical support for environmental engineers working on projects related to compliance with the Clean Air Act. Focused on refinery, utility and greenhouse gas initiatives.

University of Illinois at Chicago, Department of Environmental and Occupational Health, Chicago, IL
2009- 2010
Research Assistant
Worked with two doctoral candidates on their projects involving analyzing contaminants found in the sediment of the Great Lakes.

- Published: Ruiqiang Yang, Hua Wei, Jiehong Guo, Colin McLeod, An Li, Neil C. Sturchio**.** Historically and Currently Used Dechloranes in the Sediments of the Great Lakes**.** *Environmental Science & Technology* 2011 *45* (12), 5156-5163.

University of Ireland, Environmental Microbiology Research Department, Galway, Ireland
2007
Research Assistant
Studied anaerobic digestion and the effectiveness of methane producing bacteria in treating industrial waste using upflow anaerobic sludge blanket (UASB) reactors.

Education

Fairfield College Preparatory School, *2004* - Fairfield, CT
Lehigh University, *2008 - Bachelor of Arts in Biology*, *Minor in Economics*, Bethlehem, PA
University of Illinois at Chicago, *2009-2010* - Fulfilled masters program prerequisites, Chicago, IL
Lehigh University, *2011 - Master of Science in Environmental Engineering*, Bethlehem, PA